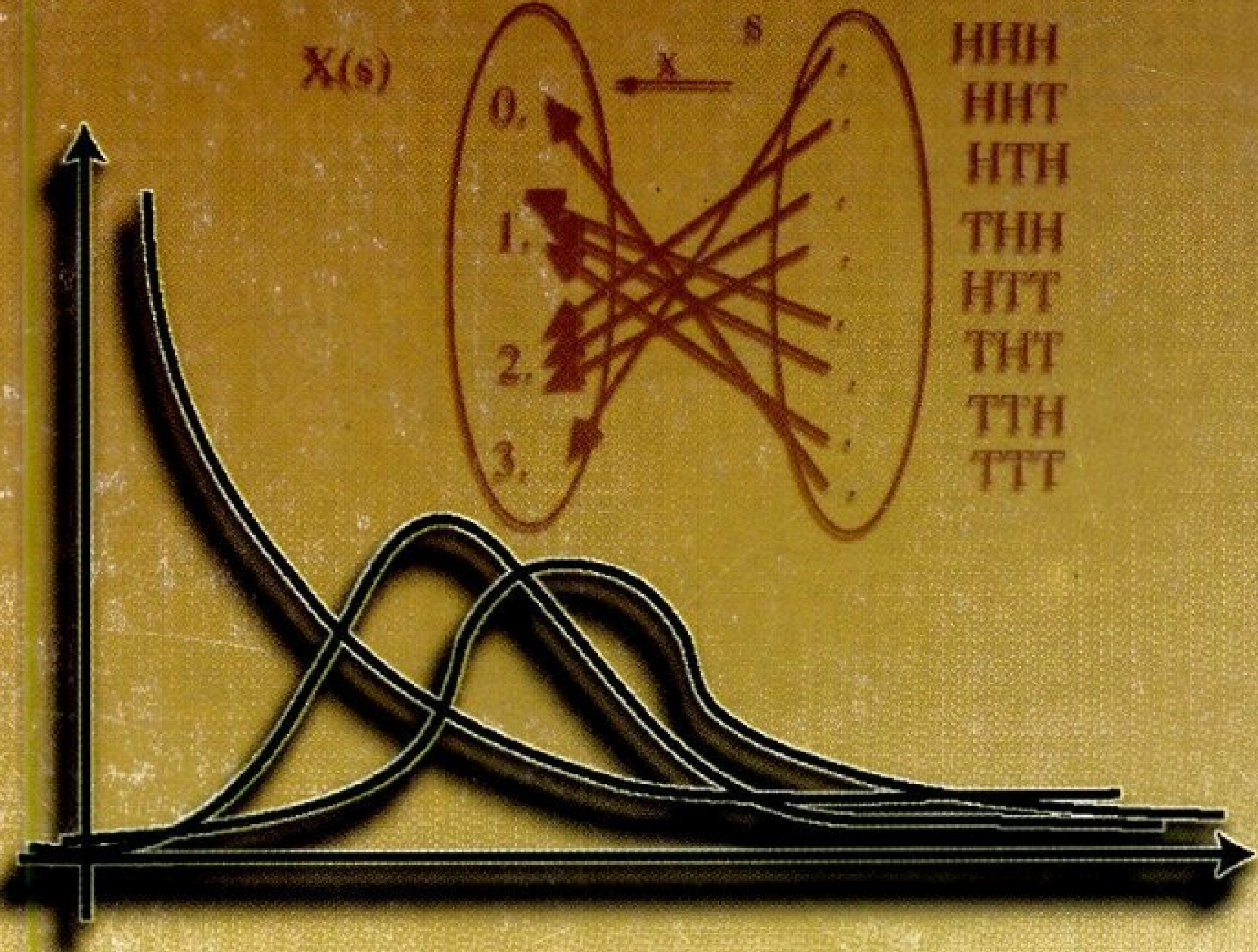
أساسيات طرق التحليل الإحصائي



تأليف

الدكتور عدنان بن ماجد عبد الرحمن بري

الدكتور محمود محمد إبراهيم هندي

الدكتور الحسيني عبد البر إبراهيم راضي

جامعة الملك سعود

النشر العلمي و المطابع

٦١١

بسم الله الرحمن الرحيم

أساسيات طرق التحليل الإحصائي

تأليف

الدكتور عدنان بن ماجد عبدالرحمن بري

الدكتور محمود محمد إبراهيم هندي

الدكتور الحسيني عبدالبر إبراهيم راضي

قسم الإحصاء، كلية العلوم، جامعة الملك سعود

النشر العلمي والمطابع - جامعة الملك سعود

ص. ب ٢٤٥٤ الرياض ١١٤٥١ - المملكة العربية السعودية

فهرسة مكتبة الملك فهد الوطنية أثناء النشر
بري، عدنان بن ماجد عبدالرحمن
أساسيات طرق التحليل الإحصائي / عدنان بن ماجد بري، محمود هندي، الحسيني عبدالبر إبراهيم راضي - الرياض.
٥٤٨ص، ١٧سم × ٢٤ سم
ردمك ٣-٦٩٥-٠٥-٩٩٦٠
١- الإحصاء التحليلي أ- هندي، محمود محمد (م. مشارك)
ب- راضي، الحسيني عبدالبر إبراهيم (م. مشارك) جـ- العنوان
ديوي ٥١٩,٥ ١٨/٣٨١٤

رقم الإيداع ١٨/٣٨١٤

تم تحكيم الكتاب بواسطة لجنة متخصصة شكلها المجلس العلمي بالجامعة، وقد وافق المجلس على نشره في اجتماعه التاسع للعام الدراسي ١٤١٥/ ١٤١٦هـ المعقود بتاريخ ٩/ ٧/ ١٤١٥هـ الموافق ١١/ ١٢/ ١٩٩٤م.

النشر العلمي والمطابع ١٤١٩هـ

تقديم

لقد أصبح الإتجاه العام، في الوقت الحاضر، في كثير من العلـوم مـثـل الفيزياء، الكيمياء، الأحياء، الإلكترونات، الاتصالات، الزراعة، التعليم، علم النفس، العلوم الطبية، العلوم الهندسية، وغيرها، هو استخدام طرق القيـاس الكمية والوصفيّة واستخدامها في التحليل الإحصائي وذلك لمواصلة تقدمها المطّرد. والهدف من هذا الكتاب هو تقديم بعض المفاهيم الاحصائية الأساسية دون الخوض في ذكر البراهين الرياضية المعقدة ، وذلك مراعاة للتباين في مستـوى الـطـلاّب المبتدئون وكذلك الباحثون غير الملمين بالرياضيات المتقدمة . ثم أوردنا كذلـك كثيرا من المفاهيم الإحصائية والعمليات اللازمة لإجرائها على شكـل خـطـوات بغرض تبسيطها ، وسهولة فهمها ومتابعتها .وجاء هذا الكتاب باللغة العربية في كل محتوياته ما عدا المعادلات فهي مطبوعة بالرموز اللاتينية، والهدف من هـذا الإجراء هو التيسير على القارىء في الاستفادة من المراجع الأجنبية المذكورة فـي آخر الكتاب عند الرجوع اليها .

ويحتوي هذا الكتاب على المقدمة والإحصاء الوصفي وهـو مـا ورد فـي الفصلين الأول والثاني . ويحتوي على مبادىء نظرية الاحتمالات والمتغـيـرات العشوائية، التوزيعات الاحتمالية المنفصلة مثل توزيع ذي الحدين، فوق الهندسي وبواسون، المتصلة مثل التوزيع الطبيعي، توزيع-t، توزيع مربع كاي وتوزيع F ، وهو ما ذكر في الفصلين الثالث والرابع .

ويحتوي كذلك على طرق التحليل الإحصائي وهو توزيع المعاينة، نظرية النهاية المركزية، التقدير، اختبارات الفروض، اختبارات مربع كاي، تصميم التجارب الإحصائية، تحليل التباين في اتجاه واحد وكذلك في اتجاهين ، الارتباط والانحدار الخطي وغير الخطي ، والاختبارات اللا معلمية، وهو ما اشتملت عليه الفصول العشرة الباقية ابتداء من الفصل الخامس حتى الفصل الرابع عشر والأخير.

كما يوجد في نهاية كل فصل عدد من الأمثلة والتمارين التي تساعد الطالب على تعميق المفاهيم والتطبيقات .

ونود أن نتقدم بالشكر الجزيل إلى الزملاء الذين أثروا المكتبة العربية بعدد غير قليل من كتب الإحصاء والتي استفدنا منها كثيرًا عند إعداد هذا الكتاب والتي أوردناها في نهاية الكتاب تحت عنوان المراجع العربية والأجنبية .

وأخيرا فإننا نتقدم بالشكر والتقدير لكل من ساهم معنا في إخراج هذا الكتاب وبشكله الراهن كما نتقدم بالشكر والتقدير إلى من يلتمس لنا العذر إن لاحظ أي نقص أو قصور فالشمول أمر صعب والكمال لله وحده .

المؤلفون

المحتويات

الصفحة

الصفحة

الفصل الثالث : مبادىء الاحتمالات

الفصل الرابع : المتغيّرات العشوائيّة والتوزيعات الاحتماليّة

الصفحة

الصفحة

الفصل السابع : اختبارات الفروض الإحصائية

الصفحة

الفصل الثامن : اختبارات مربع كاي

الفصل التاسع : أساسيات تصميم التجارب

الصفحة

الصفحة

الفصل العاشر : تحليل التباين في اتجاه واحد

الفصل الحادي عشر : طرق المقارنات المتعددة .

الصفحة

الصفحة

الفصل الرابع عشر : الاختبارات اللامعلمية

الجــداول

الصفحة

الفصل الأول

مقدمة عن الإحصاء
Introduction to Statistics

● مقدمة ● تاريخ علم الإحصاء ● تعريف علم الإحصاء ● بعض المصطلحات الأساسية في علم الإحصاء ● المراحل الأساسية للبحث الإحصائي ● الأخطاء الإحصائية ● تمارين

(١ , ١) مقدمة
Introduction

يعتبر اللفظ إحصاء من الألفاظ القديمة والكلمة الإنجليزية (Statistics) مشتقة من الكلمة اللاتينية (Status) وهي تعني الدولة . لقد عرفت الإحصاءات (التعدادات) من قديم الأزمنة حيث كانت تستخدم لأغراض حربية وضريبية حتى تتمكن كل دولة من تكوين جيش قوي يستطيع الدفاع عن حدودها ، وكذلك حتى تتمكن الدولة من فرض الضرائب وتجميع الأموال اللازمة لتمويل الجيش وإدارة شئون البلاد . ثم توسعت عمليات التعداد والإحصاءات لتشمل بيانات المواليد والوفيات، والإنتاج والاستهلاك . وفي صدر الإسلام كان الجيش المسلم يقوم بعمل إجراءات لمعرفة عدد أفراد جيش الأعداء عن طريق إحصاء عدد مواقع نيرانهم لطهي الطعام أثناء الليل وبذلك يتمكن من إحصاء جنود الأعداء لكل موقع . وزادت أهمية الإحصاءات في القرن الثامن عشر وخاصة بعد ظهور الثورة الصناعية حينما أيقن رجال الأعمال بضرورتها من أجل اتخاذ قرارات سليمة .

ويخدم الإحصاء الباحثين في جميع الميادين العلمية ومتخذي القرارات في المجالات المختلفة سواء كانت إحيائية (بيولوجية)، هندسية ، طبية ،

اقتصادية . . .الخ . وعلى سبيل المثال فإن الباحث في مجال الاقتصاد يستطيع أن يختبر نظرياته عن سلوك المستهلك أو علاقة المستخدم والمنتج عن طريق الطـرق الإحصائية . كما أن الباحث في مجال الصيدلة يستخدم الأساليب نفسها لقياس كفاءة دواء جديد ، وفي مجال الطب لايجاد العلاقة بين التدخين ومرض معين ، وفي مجال الزراعة لمعرفة آثار الأسمدة المختلفة على محصول معين . كما يستخدمها الباحث في مجال التخطيط في تقويم مشاكل الصناعة وعمليات ضـبـط جـودة الإنتاج وموثوقية صناعات كثيرة مثل الأدوات الكهربائية . . . الخ . نستعرض في هذا الفصل بعض المعلومات التاريخية عن علم الإحصاء وتعريفه وتعريف وبعض مصطلحاته الأساسية ثم المراحل الأساسية للبحث الإحصائي ومن ثم الأخطاء الإحصائيةالتي قد تحدث في أي عملية إحصائية.

(٢ , ١) تاريخ علم الإحصاء

History of Statistical Science

إن دراسة تاريخ أي علم تلقي الضوء عن كيفية تطوره. ولقد نشأ الإحصاء كدراسة لوصف الدولة، خاصة أمورها الاقتصادية وخصائصها الديموجرافيـة . وفي التاريخ القديم، أصدر القيصر أوجيسيس أمراً بفرض ضريبة على كل فرد في مملكته ، استدعى ذلك أن يسجل كل شخص اسمه عند أقرب إحصائي (في ذلك الوقت كان جابي الضريبة). كما أمر وليم الفاتح ملك انجلترا بمسح شامل لجميع الأراضي في إنجلترا وذلك لأغراض عسكرية ثم عمل كتاب سمي domesday .

بعد ذلك، وفي القرن السابع عشر، استخدمت تطبيقات الاحتمال التجريبي في التأمين على السفن . وظهرت دراسات السكان ودراسات اختلاف القياسات وجدولتها وتم تطوير البيانات السكانية عن حالات الزواج، المواليد، الوفيات، العمر، التعليم . . إلخ . وبدراسة سلاسل زمنية عن تلك البيانات لوحظ وجود قفزات عشوائية ، أدى ذلك إلى ظهور ما يسمى بالأخطاء العشوائية وتـوزيـع الخـطـأ وظهرت نماذج الصدفة وتمكن دي موافر (1754 - 1667) من استنباط معادلة المنحنى

الطبيعـي (1733) وكان لا يدرك أهميتها حتى وجدها كارل بيـرسـون (1924) في احدى المكتبات . كما توصل إلى نفس النتائـج كـل من لابـلاس (1827 - 1749) وجاوس (1858 - 1777) - كل منهما على حدة - كما توصل جاوس وآخرون إلى إثبات وجود بعض النجوم باستخدام الإحصاء .

إلى جانب ارتباط تطور علم الإحصاء بالدراسات الأكتوارية والديموجرافية والفلك، فقد أسهمت جميع العلوم مثل علم النفس والطب والوراثة والعـلـوم الأحيائية (البيولوجية) والزراعية في تطوره أيضًا. فمثلا توزيع بواسون وعلاقتـه بالإصابات الناجمة عن رفس الخيل وكذلك علاقته بعدد كرات الدم الحمراء . . . إلخ. ساعد الإحصاء أيضا مندل (1884 - 1822) في دراسته لعلم الأجنة .

ومن الذين أسهموا مساهمة فعالة في تطور علم الإحصاء كل من:

كارل بيرسون (1936 - 1857): وهو عالم إنجليزي فيزيائي تحول إلى دراسة العلـوم الإحصائية وله إسهامات كثيرة في علم الإحصاء نذكر منها جمع وتلـخـيـص وجدولة بيانات تجريبية كثيرة، وتعريف معامل الارتباط ومعامل الارتباط الجزئي وتقديره واستخدم حالة الإمكان في حالات متعددة وأيضا اختبار مربع كاي (χ^2) واستخدامه لاختبارات جودة التوفيق وتحليل جداول التوافق . . . . .

سير رونالد فيشر (1962 - 1890): وهو عالم إنجليزي يعتبر من الذين أضافوا الكثير لعلم الإحصاء وقد استخدم الإحصاء بكثرة في الزراعة والعلوم الأحيائية وعلـم الأجنة وذلك من خلال عمله في محطة روثمند للتجارب بإنجلترا . وهو الـذي أجاز اختبار t الذي وضع بواسطة جوست ولكن بعد تحويله باقتراح فيشر سمي اختبار ستودنـت t ، ويستخدم اختبـار t للعينات الصغيرة . كما وضع فيـشـر أساسيات علم تصميم التجارب وتحليل التباين واختبار F وكثير من الإسهامات في نظرية التقدير . . . .

نيمان (1981 - 1894) وبيرسون (1980 - 1885): قدما نظرية اختبار الفرضيات الإحصائية وما تبع ذلك من تأثير على التطبيقات العملية . . . .

ويقسم بعض الإحصائيين تطور علم الإحصاء إلى فترات هي :

الأولى: من 1920 - 1890 : هي فترة كارل بيرسون .

الثانية : من 1936 - 1921 : هي فترة رونالد فيشر .

الثالثة : من 1949 - 1937 : هي فترة نيمان وبيرسون .

الرابعة : من 1950 حتى وقتنا الحاضر : أصبحت الإحصاء أكثر استخداما للرياضيات بجميع فروعها وأصبحت جزءًا من علم الإحصاء الرياضي . كما أن ظهور الحاسبات الإلكترونية وظهور حزم البرامج الجاهزة سهلت استخدام الطرق الإحصائية لكثير من الباحثين .

(٣,١) تعريف علم الإحصاء

Defintion of Statistical Science

يعرف علم الإحصاء بأنه ذلك الفرع من العلوم الذي يختص بالطرق العلمية لجمع البيانات، تنظيمها، تلخيصها، عرضها وتحليلها وذلك للوصول إلى نتائج مقبولة تؤدي إلى اتخاذ قرارات سليمة .

وعلى الرغم من شمول هذا التعريف على كثير من فروع علم الإحصاء إلا أنه لم يأخذ في اعتباره عدم التأكد والعشوائية وهي أهم ما يميز علم الإحصاء . لأن علم الإحصاء يتعامل مع مشاكل يكون قانون التأثر والسبب (cause and effect) للظاهرة المشاهدة غير معروف للباحث ويحتاج الأمر اسلوبا موضوعيا . ويصبح التعريف المعقول، وللأسباب السابقة، لعلم الإحصاء : أنه علم ابتكار وتطوير وتطبيق أساليب تمكننا من تقويم عدم التأكد والعشوائية للاستدلال الإحصائي . ويعتبر علم الإحصاء لكثير من الباحثين المنطق السليم المدعم بمزيج قوي من

الأساليب الرياضية .ويقسم البعض علم الإحصاء ومن وجهة نظر تطبيقية إلى ثلاثة أقسام:

أولا: الإحصاء الوصفي

ويختص الإحصاء الوصفي (descriptive statistics) بتلخيص البيانات المستخدمة للبحث الإحصائي في جداول إحصائية أو رسوم بيانية أو حساب مقاييس إحصائية مثل النزعة المركزية والتشتت والالتواء والتفلطح وقياس قوة الارتباط بين ظاهرة وأخرى .

ثانيا: الاستدلال الإحصائي

ويختص الاستدلال الإحصائي (statistical inference) باستخلاص نتائج عامة من بعض البيانات حيث يتم ذلك باستخدام المعاينة الإحصائية (statistical sampling). ويجدرالإشارة هنا إلى المجتمع الإحصائي (statistical population) وهو عبارة عن جميع المفردات موضع الدراسة والتي نرغب معرفة حقائق عنها سواء كانت على شكل إنسان أو حيوان أو جماد أو درجات امتحان أو منازل أو مزارع أو متاجر أو سفن . . . إلخ . فعلى سبيل المثال، قد يكون المجتمع المطلوب دراسته هو سكان مدينة معينة أو منازل هذه المدينة أو جميع درجات امتحان مادة معينة . . . الخ . أما العينة الإحصائية فهي جزء من المجتمع محل الدراسة تحقق شروط وخصائص المجتمع وتكون صغيرة يسهل دراستها عكس المجتمع قد يكون كبيراً ويصعب دراسته . وتمثل العينة على سبيل المثال جزء من سكان مدينة معينة أو جزء من منازل هذه المدينة أو جزء من درجات طلاب أحد المواد الدراسية وهكذا . ويوجد علم خاص بطرق أخذ العينات يسمى المعاينة الإحصائية . ومن المعاينة الإحصائية يتم الوصول إلى نتائج يمكن تعميمها على المجتمع الإحصائي محل الدراسةككل. وفي حالة احتمال عدم تمثيل العينة للمجتمع تمثيلا حقيقيا فإن الاستدلال الإحصائي يمكن الباحث الإحصائي من قياس الخطأ الناتج عن ذلك.

ثالثا: التنبؤ

يختص التنبؤ (predection) بتحديد ما سيحدث للظاهرة محل البحث في فترة زمنية مقبلة . فإذا كان لدينا على سبيل المثال علاقة بين المبيعات من سلعة ولتكن Y والزمن بالسنوات وليكن X فإنه يمكن استخدام العلاقة بين Y و X في التنبؤ بمبيعات هذه السلعة في فترة زمنية مقبلة استنادا إلى استمرار العلاقة في المستقبل على ماكانت عليه في الماضي .

(٤ ,١) بعض المصطلحات الأساسية في علم الإحصاء

نقدم في هذا البند بعض المصطلحات المستخدمة في علم الإحصاء ونذكر منها :

المجتمع

يمثل المجتمع (population) جميع المفردات محل الدراسة والتي لها خصائص مشتركة .فقد يكون المجتمع محدودا مثل جميع طلاب جامعة الملك سعود في العام الدراسي الحالي ، أو قد يكون غير محدد مثل جميع الأسماك الموجودة في الخليج العربي . . الخ .

العيّنة

تمثل العينة (sample) جزءاً من مفردات المجتمع محل الدراسة ويشترط في العينة بأن تكون ممثلة لجميع صفات المجتمع محل الدراسة والتي اختيرت منه مثل مجموعة من طلاب إحدى كليات جامعة الملك سعود أو مجموعة أسماك مصطادة من الخليج العربي خلال أحد الأيام . . . الخ . تستخدم البحوث الإحصائية العينة للحصول على نتائج سريعة مقارنة بأسلوب الحصر الشامل . ويجب أن تختار العينة حسب أسلوب المعاينة .

المتغير

ويمثل المتغير (variable) قيمًا لظاهرة ما حيث تتغير قيم مفردات الظاهرة من مفردة إلى أخرى مثل ظاهرة الطول والوزن والعمر لطلاب جامعة الملك سعود ويرمز للمتغير بالرموز الكبيرة .. $X, Y, Z,$ ، وقيم البيانات لكل متغير يرمز لها بالرموز الصغيرة .. $x, y, z,$. فمثلا متغير الطول لطلاب جامعة الملك سعود X لعينة حجمها n طالب أطوالهم هي $x_1, x_2, \ldots, x_n$ أي بيانات أطوال الطلاب المختلفة لمتغير الطول X للعينة n محل الدراسة . ويمكن أن تكون بيانات المتغير شاملة لأفراد المجتمع بأسره .

وقد يكون المتغير كمي مثل الطول والوزن والعمر . . . الخ . وقد يكون المتغير وصفياً مثل لون العيون والبشرة ولون الشعر . . . الخ .

المعلمة

تمثل المعلمة (parameter) مؤشر يصف المجتمع، مثل المتوسط لأطوال طلاب جامعة الملك سعود في العام الحالي ويرمز له بالرمز μ وكذلك تباين هؤلاء الطلاب σ^2 ونسبة الناجحين منهم P.

الإحصاءة

تمثل الإحصاءة (statistic) مؤشرًا يحسب من العينة ويستخدم تقديرا لمعلمة المجتمع محل الدراسة، فمثلا متوسط أطوال عينة من طلاب جامعة الملك سعود يرمز له بالرمز $\overline{x}$ يمثل تقديرا لمعلمة المجتمع μ وكذلك تباين العينة s^2 تقديرا لمعلمة المجتمع σ^2 وكذلك نسبة الناجحين في العينة r تقديرا لنسبة الناجحين لمعلمة المجتمع P . . . الخ .

(١,٥) المراحل الأساسية للبحث الإحصائي

سنذكر فيما يلي بعض المراحل الاساسية التي يمر بها البحث الإحصائي :

تحديد المشكلة محل البحث

يجب على الباحث أن يحدد المشكلة في شكل أسئلة محددة . وتحديد المشكلة بهذه الصورة يرشد الباحث إلى البيانات الواجب جمعها بالإضافة إلى الطرق التي سيتبعها لحل هذه المشكلة .

مرحلة جمع البيانات الخاصة بالبحث

بعد تحديد المشكلة موضع البحث تكون المرحلة التالية هي جمع البيانات الخاصة بها . لذا يجب معرفة ماهي مصادر البيانات، وهل سبق جمع مثل هذه البيانات في بحوث سابقة، حتى يمكن الاستفادة منها بحيث لا يضيع الوقت والمجهود في إعادة جمعها . والمعروف أنه يوجد مصدران رئيسان لجمع البيانات هما : المصدر التاريخي وهو يؤخذ عادة من السجلات مثل سجلات المواليد والوفيات وإحصاءات هيئة الأمم وغيرها . والمصدر الثاني ميداني حيث تجمع البيانات من المجتمع مباشرة إما بالاتصال المباشر أو بالبريد أو بالهاتف .

وفي حالة جمع البيانات تصمم استمارة إحصائية (questionnaire) تتكون من صفحة أو أكثر تتضمن مجموعة من الأسئلة المطلوب الإجابة عليها من المفردات محل الدراسة في العينة أو المجتمع محل الدراسة . ومن خصائص الأسئلة يجب أن تكون واضحة ومباشرة ويسهل الإجابة عليها ، وتكون كافية وتفى بأغراض الدراسة المطلوبة بحيث لا تكون مختصرة فتحذف أي معلومات أساسية أو مفصلة أكثر من اللازم تحتوي على تفصيلات غير مطلوبة . تملأ هذه الاستمارة من مصادر جمع البيانات سواء كانت تاريخية أو ميدانية .

القيام بأبحاث ميدانية

بعد إجراء المرحلتين السابقتين قد يتضح أن البيانات المتوافرة لا تمثل كل البيانات اللازمة للبحث . لذا فإن الخطوة التالية هي القيام بتجميع البيانات الناقصة من مصادرها الأولية إما بالاتصال المباشر أو غير المباشر مثل البريد والهاتف وغيرهما .

تصنيف البيانات

يقصد بتصنيف أو تبويب البيانات هو وضع البيانات المتشابهة في فئات تشترك في خاصية، أو وحدة معينة ووضع البيانات المتشابهة في فئات تشترك في خاصية تميزها عن باقي بيانات الفئات الأخرى. وتعتبر مرحلة التصنيف هذه هي الخطوة الأولى في عملية تحليل البيانات .

عرض البيانات

بعد تصنيف البيانات تأتي مرحلة عرضها في شكل جداول تكرارية أو علي شكل رسوم بيانية . . الخ .

التحليل الإحصائي

يكتفي الباحث الإحصائي أو متخذ القرار في بعض الأحيان بالجداول التكرارية أو الرسوم البيانية . إلا إنه في بعض الأحيان قد يكون مطلوب عمل تحليل إحصائي حتى يصل الباحث إلى النتائج المطلوبة التي تساعد على اتخاذ قرار. وتعتبر مرحلة التحليل الإحصائي هي الخطوة الأخيرة في مراحل البحث الإحصائي.

(١,٦) الأخطاء الإحصائية

تنتج بعض الأخطاء الإحصائية إما عن خطأ في تدوين المعلومات أو خطأ في عملية الحسابات مما يؤدي للوصول إلى نتائج خاطئة . بالإضافة إلى نوع آخر من الأخطاء هو سوء تحليل البيانات الإحصائية. يمثل النوع الأخير من الأخطاء

كل من التحيز وعدم قابلية البيانات للمقارنة والتنبؤ غير السليم للاتجاه العام ووضع مسببات خاطئة والمقارنة بحالة غير عادية وأن تكون العينة غير ممثلة للمجتمع. ونتناول كل فرع من هذه الأخطاء بشيء من التفصيل فيما يلي :

التحيّز

التحيز (bias) هو في أن يعطي الشخص أهمية أكثر أو أقل للمعلومات من وجهة نظره عن تلك التي تعطيها البيانات والقياسات الحقيقية . وهناك حالة قصوى للتحيز عندما تكون النتيجة محددة مسبقا .

عدم قابلية البيانات للمقارنة

في بعض الأحيان، يتطلب من الباحث عمل مقارنة لمتغير معين وتكون البيانات ذاتها غير قابلة للمقارنة ، مثلا عند الحاجة لعمل مقارنة لعدد الوفيات الناتجة من مرض معين ، فقد تظهر السنوات الأخيرة معدل متزايد لهذه الظاهرة ناتج من أن تحديد سبب الوفاة أصبح أكثر دقة عما كان عليه في الماضي وبالتالي فلا يمكن إجراء المقارنة في هذه الحالة .

افتراضات خاطئة خاصة بالعلاقة السببية

من الأخطاء الشائعة استنتاج حدوث ظاهرتين في آن واحد وهذا قد لا يعني أن أحدهما مسبب للآخر وربما في الحقيقية أن كليهما حدثا نتيجة مسبب ثالث . مثلا تغيرات حجم النشاط الاقتصادي تعتمد على سعر الفائدة. حيث ينكمش النشاط الاقتصادي عقب رفع سعر الفائدة، وينتعش هذا النشاط عقب خفض سعر الفائدة . إلا إن هناك من الدلائل على أن العلاقة المسببة هذه لا تفسر ظاهرة الدورة التجارية تفسيرا كاملا أي أنه توجد مسببات أخرى .

المقارنة بأساس غير عادي

تتطلب مقارنة البيانات الإحصائية فترة عادية. فإذا كانت الفترة المتخذة أساساً للمقارنة فترة غير عادية، فإن المقارنات تؤدي إلى نتائج مضللة. فمثلا إذا أردنا مقارنة محصول القمح بالمملكة العربية السعودية في سنة معينة بإحدى السنوات السابقة فيجب التأكد من أن السنة المتخذة أساسا للمقارنة هي إحدى السنوات التي يكون المحصول فيها مصابا نتيجة ظروف جوية غير عادية. حيث إن المقارنة في هذه الحالة سوف تظهر تحسنا وهميا كبيرا في إنتاج القمح.

عدم سلامة العيّنة

يعتمد التحليل الإحصائي بدرجة كبيرة على استخلاص النتائج من العينات لتحديد خصائص المجتمع. ويجب أن تكون العينة ممثلة للمجتمع. وعندما تختار العينة وتكون غير ممثلة للمجتمع فإن النتائج تكون خاطئة ولا تعكس خصائص المجتمع.

(١,٧) تمارين

١ - عرف علم الإحصاء ؟

٢ - اذكر أمثلة عن المجتمع الإحصائي والعينة الإحصائية.

٣ - اذكر ما تعرفه عن تاريخ علم الإحصاء وتطوره.

٤ - اذكر ما تعرفه عن الأخطاء الإحصائية مع إعطاء بعض الأمثلة.

٥ - اذكر ما تعرفه عن :
المعلمة - الإحصاءة - المتغير - البيانات - مصادر جمع البيانات - الاستمارة الإحصائية.

٦ - اذكر بعض الخطوات المهمة التي يجب أن يتبعها الباحث عند إجراء بحث يستخدم فيه الأسلوب الإحصائي.

الفصل الثاني

طرق عرض البيانات الإحصائية ووصفها
Presentation and Description of Statistical Data

● مقدمة ● عرض البيانات جدوليا ● عرض البيانات بيانيـا ● مقاييس النزعة المركـزيـة ● الربيعات و العشيرات والمئينات ● مقاييس التشتت ● مقاييس الالتواء والتفلطح ● تمارين

(٢ , ١) مقدمة

إن الخطوة التالية لجمع البيانات هي تبويب هذه البيانات ، فالبيانات الأولية تكون غير منتظمة ، فيجب ترتيبها وتصنيفها في شكل جدول لتسهيل تفسيرهـا وتحليلها. وكثيرا ما تتم هذه العملية بطريقة آلية باستخدام الحاسب الآلي. ويجب أن يتم تبويب البيانات بواسطة الحاسب في حالة وجود عدد كبير جداً من البيانات . أما إذا كان عدد البيانات بسيطا نسبيا فالتبويب في هذه الحالة يتم يدويا وفيمـا يلي سنتناول بالشرح طرق التبويب اليدوي . ثم نشرح طرق عرض هذه الجداول بيانيا ثم حساب مقاييس النزعة المركزية وكذلك حساب الربيعيات والعشـيـرات والمئينات وكذلك حساب مقاييس التشتت ومقاييس الالتواء والتفلطح ونـوضـح ذلك بالشرح والتفصيل فيما يلي.

(٢ , ٢) عرض البيانات جدوليًا (Tabulation)

تنظم البيانات الأولية المأخوذة لمفردات عينة من المجتمع محل الدراسـة أو للمجتمع نفسه إذا كان محدودا في جداول تكرارية لامكانية الاستفادة من هـذه

البيانات . واستخلاص المؤشرات التي تصف هذه البيانات سواء كانت هذه البيانات وصفية أو كمية وسنتناول بالشرح والتفصيل تلخيص البيانات في جداول تكرارية في الحالات الثلاث التالية :

أولا : البيانات الوصفية .

ثانيا : البيانات الكمية ذات المدى الصغير نسبيا .

ثالثا : البيانات الكمية ذات المدى الكبير نسبيا .

أولا: البيانات الوصفية

تلخص البيانات الإحصائية الوصفية (qualitative data) أولا في جدول يسمى جدول تفريغ البيانات الإحصائية وهو يتكون من ثلاث خانات . الخانة الأولى أو العمود الأول تكتب فيه الصفة للبيانات الوصفية. في الخانة الثانية توضح العلامات وهي عبارة عن حزم كل حزمة مكونة من خمسة خطوط، أربعة رأسية والخامس مائل يحزم الخطوط الأربعة الرأسية وبذلك تصبح الحزمة على الصورة (~~||||~~). يكتب في الخانة الثالثة والأخيرة مجموع العلامات أمام كل صفة ويسمى تكرارا لهذه الصفة ونوضح ذلك بالمثال التالي.

مثال (١ , ٢)

في إحدى المؤسسات كانت بيانات الحالة الاجتماعية لعدد ٣٠ عاملا هي: متزوج، أعزب، متزوج، متزوج، مطلق، متزوج، أرمل، متزوج، متزوج، متزوج، أعزب، أعزب، أعزب، مطلق، متزوج، أعزب، أعزب، متزوج، مطلق، متزوج، أرمل، أعزب، أعزب، أعزب، متزوج، متزوج، متزوج، متزوج، متزوج، متزوج . كون جدولا لتفريغ البيانات ثم جدولا تكراريا يوضح الحالة الإجتماعية للعمال .

الحل

نكون جدول تفريغ البيانات رقم (١ , ٢) ثم الجدول التكراري رقم (٢ , ٢) كما يلي :

جدول (٢ , ٢) التكراري

عدد العمال	الحالة الإجتماعية
16	متزوج
9	أعزب
3	مطلق
2	أرمل
30	المجموع

جدول (١ , ٢)

الصفة	العلامات	التكرار
متزوج		16
أعزب		9
مطلق		3
أرمل		2
	المجموع	30

ثانيًا: البيانات الكمية ذات المدى الصغير نسبيا

تلخص البيانات الكمية (quantitative data) وذات المدى الصغير نسبيا مثل ما تم في البيانات الوصفية من تكوين جدولا لتفريغ البيانات ثم جدولا تكراريا ونوضح ذلك بالمثال التالي.

مثال (٢ , ٢)

في دراسة اجتماعية كان حجم 25 أسرة كما يلي :

2,2,3,3,3,4,5,3,4,4,3,5,3,3,3,5,4,4,4,3,2,2,2,3,2.

كون جدول لتفريغ البيانات ثم جدولا تكراريا يوضح توزيع حجم الأسرة.

الحل

نكون جدول تفريغ البيانات (٣ , ٢) ثم الجدول التكراري (٤ , ٢) كما يلي:

جدول (٢,٤)

عدد الأسر	حجم الأسرة
6	2
10	3
6	4
2	5
25	المجموع

جدول (٢,٣)

التكرار	العلامات	الصفة
6	卌 I	2
10	卌 卌	3
6	卌 I	4
2	III	5
25		المجموع

ثالثا: البيانات الكمية ذات المدى الكبير نسبيا

في هذه الحالة تقسم البيانات إلى فئات، تقسم قيم البيانات إلى فئات متساوية الطول. تفرغ البيانات على الفئات لنحصل على تكرار لكل فئة . ولشرح خطوات عمل جدولا تكراريا وطريقة تكوين فئاته لاحظ حل المثال التالي.

مثال (٢,٣)

البيانات التالية درجات 30 طالبا في أحد المواد الدراسية :

51	59	70	74	73	90
91	72	83	89	50	80
62	82	87	76	91	76
71	96	81	88	64	82
80	81	75	85	74	90

كون جدولا تكراريا لدرجات الطلاب .

الحل

نكون جدول تفريغ البيانات (٢,٥) من ثلاث خانات الخانة الأولى تمثل فئات الدرجات والخانة الثانية العلامات والخانة الثالثة تمثل التكرار (عدد الطلاب)

ثم نلخص جدول تفريغ البيانات (٢,٥) في جدول تكراري من خانتين الخانة الأولى فئات الدرجات والثانية التكرار (عدد الطلاب) كما في جدول (٢,٦) وفيما يلي خطوات عمل الفئات للجدول التكراري (٢,٦) .

أ) نحسب طول المدى للبيانات R وهو عبارة عن الفرق بين أكبر قراءة في البيانات وبين أصغر قراءة أي أن :

$R = 96 - 50 = 46$ درجة

ب) نختار عدد الفئات k وليكن $k = 5$ مثلا .

ج) نحسب طول الفئة L وذلك بقسمة المدى R على عدد الفئات k ثم يقرب الناتج لأقرب وحدة دقة وفي مثالنا هذا وحدة الدقة هي واحد درجة ويكون طول الفئة كما يلي :

$$L = R \div k$$

$$= 46 \div 5 = 9.4$$ درجة

ويقرب طول الفئة L ليصبح $L = 10$ درجات .

إذا كانت البيانات تحتوي على مشاهدات كسرية يقرب طول الفئة للرقم العشري المعطى في هذه البيانات .

د) يختار أصغر قراءة في البيانات بداية للفئة الأولى وفي مثالنا يكون بداية الفئة الأولى المقربة 50 درجة ثم نضيف طول الفئة على بداية الفئة الأولى لنحصل علي بداية الفئة الثانية وهو $60 = 10 + 50$ درجة ثم هكذا لبداية باقي الفئات .

هـ) لايجاد نهاية أي فئة نطرح وحدة دقة من بداية أي فئة وفي مثالنا (واحد صحيح) لنحصل على نهاية الفئة السابقة لها فمثلا نهاية الفئة الأولى = بداية

الفئة الثانية مطروحا منها واحد أي أن نهاية الفئة الأولى =1 - 60 = 59 درجة وهكذا لباقي الفئات. ثم نكون جدول تفريع البيانات (٥, ٢) والجدول التكراري (٦, ٢) لدرجات الطلاب كما يلي :

جدول (٦, ٢)

عدد الطلاب	فئات الدرجات التكرار
2	50 - 59
2	60 - 69
9	70 - 79
11	80 - 89
6	90 - 99
30	المجموع

جدول (٥, ٢)

عدد الطلاب	العلامات	الفئات
2	ǀǀ	50 - 59
2	ǀǀ	69 - 69
9	卌 ǀǀǀǀ	70 - 79
11	卌 卌 ǀ	80 - 89
6	卌 ǀ	90 - 99
30		المجموع

و) ويمكن تكوين الحدود الحقيقية (الفعلية) للفئات وذلك بطرح نصف وحدة دقة من بداية الفئة المقربة في جدول (٦, ٢) السابق لنحصل على بداية الفئة الحقيقية وفي مثالنا بداية الفئة الأولى الحقيقية هي 0.5 - 50 = 49.5 درجة وهكذا لباقي الفئات وبإضافة نصف وحدة دقة الى نهاية الفئة المقربة في جدول (٦, ٢) السابق لنحصل على نهاية الفئة الحقيقية وفي مثالنا نهاية الفئة الأولى الحقيقية هي 0.5 + 59 = 59.5 درجة.

ز) لحساب التكرار النسبي للفئات يتم بقسمة التكرار الفعلى لكل فئة على مجموع التكرارات لنحصل على التكرار النسبي لهذه الفئة ثم نحصل على التكرار المئوي بضرب التكرار النسبي في 100 لنحصل على التكرار المئوي لكل فئة .

ح) لحساب مركز الفئة يتم بجمع الحد الأدنى والحد الأعلى للفئة ثم القسمة على 2 لنحصل على مركز الفئة وفي مثالنا مركز الفئة الأولى هو = $\frac{50 + 59}{2} = \frac{49.5 + 59.5}{2} = \frac{109}{2} = 54.5$ درجة سواء كانت حدود مقربة أو حدود

حقيقية للفئة ويمكن تلخيص الخطوات السابقة في جدول (٢,٧) كما يلي :

جدول (٢,٧)

تكرار مئوي	تكرار نسبي	تكرار فعلي	مراكز الفئات	حدود حقيقية	حدود مقربة
7	0.07	2	54.5	49.5 - 59.5	50 - 59
7	0.07	2	64.5	59.5 - 69.5	60 - 69
30	0.30	9	74.5	69.5 - 79.5	70 - 79
36	0.36	11	84.5	79.5 - 89.5	80 - 89
20	0.20	6	94.5	89.5 - 99.5	90 - 99
100	1	30	المجموع		

ط) يمكن حساب الجدول التكراري المتجمع الصاعد من الجدول التكراري (٢,٤) أو الجدول التكراري (٢,٧) باستخدام حدود الفئات الحقيقية الدنيا وكتابة أقل من أمام كل منها ثم تجميع التكرارات سواء كانت فعلية أو نسبية أو مئوية لنحصل على جدول (٢,٨) حيث التجميع للتكرارات الفعلية يسمى الجدول التكراري المتجمع الفعلي وإذا كان التجميع للتكرارات النسبية يسمى بالجدول التكراري المتجمع النسبي وكذلك بالنسبة للتكرار المئوي فمثلا عدد الطلاب الحاصلين على 79.5 درجة فأقل هو $2 + 2 + 9 = 13$ طالباً بنسبة $\frac{13}{30} = 0.43$

وهكذا لباقي الفئات كما هو موضح في جدول (٢,٨) كما يلي :

جدول (٢,٨)

التكرار المتجمع المئوي	التكرار المتجمع النسبي	التكرار المتجمع الفعلي	حدود الفئات الدنيا
00	0.00	0	أقل من 49.5
07	0.07	2	أقل من 59.5
14	0.14	4	أقل من 69.5
43	0.43	13	أقل من 79.5
80	0.80	24	أقل من 89.5
100	1.00	30	أقل من 99.5

وبالمثل يمكن صياغة الجدول التكراري المتجمع الهابط (٩ , ٢) مثل جدول (٨ , ٢) واستبدال عبارة أقل من بعبارة أكثر من فمثلا عدد الطلاب الحاصلين على 79.5 درجة فأكثر هو 11 + 6 = 17 درجة بنسبة 0.57 . نلخص ذلك في الجدول (٩ , ٢) كما يلي:

جدول (٩ , ٢)

التكرار المتجمع المئوي	التكرار المتجمع النسبي	التكرار المتجمع الفعلي	حدود الفئات الدنيا
100	1.0	30	أكثر من 49.5
93	0.93	28	أكثر من 59.5
87	0.87	26	أكثر من 79.5
57	0.57	17	أكثر من 79.5
20	0.20	6	أكثر من 89.5
00	00	0	أكثر من 99.5

(٣ , ٢) عرض البيانات بيانيا

(Graphical Representation)

التمثيل بالأعمدة البيانية والخط البياني

يمكن تمثيل البيانات في الجداول التكرارية السابقة في صورة رسومات بيانية لتبسيطها وتسهيل الرؤية بين المتغيرات بأكثر سهولة من الجداول التكرارية من حيث الزيادة والنقصان للظواهر محل الدراسة ومن أهم هذه الطرق التمثيل بالأعمدة أو الخط البياني أوالرسوم الدائرية وسوف نتناول على سبيل المثال التمثيل بالأعمدة البيانية والخط البياني للحالة الاجتماعية في مثال (١ , ٢) السابق .

مثال (٤ , ٢)

ا) مثل البيانات للحالة الاجتماعية لعمال إحدى المؤسسات في مثال (١ , ٢)

السابق وذلك باستخدام الأعمدة البيانية.

ب) استخدم الخط البياني لتمثيل عدد المدارس الثانوية للذكور والإناث في المملكة العربية السعودية في الفترة من ١٣٩٥/٩٦ إلى ١٤٠٠/١٤٠١. وذلك باستخدام البيانات المعطاة بالجدول التالي

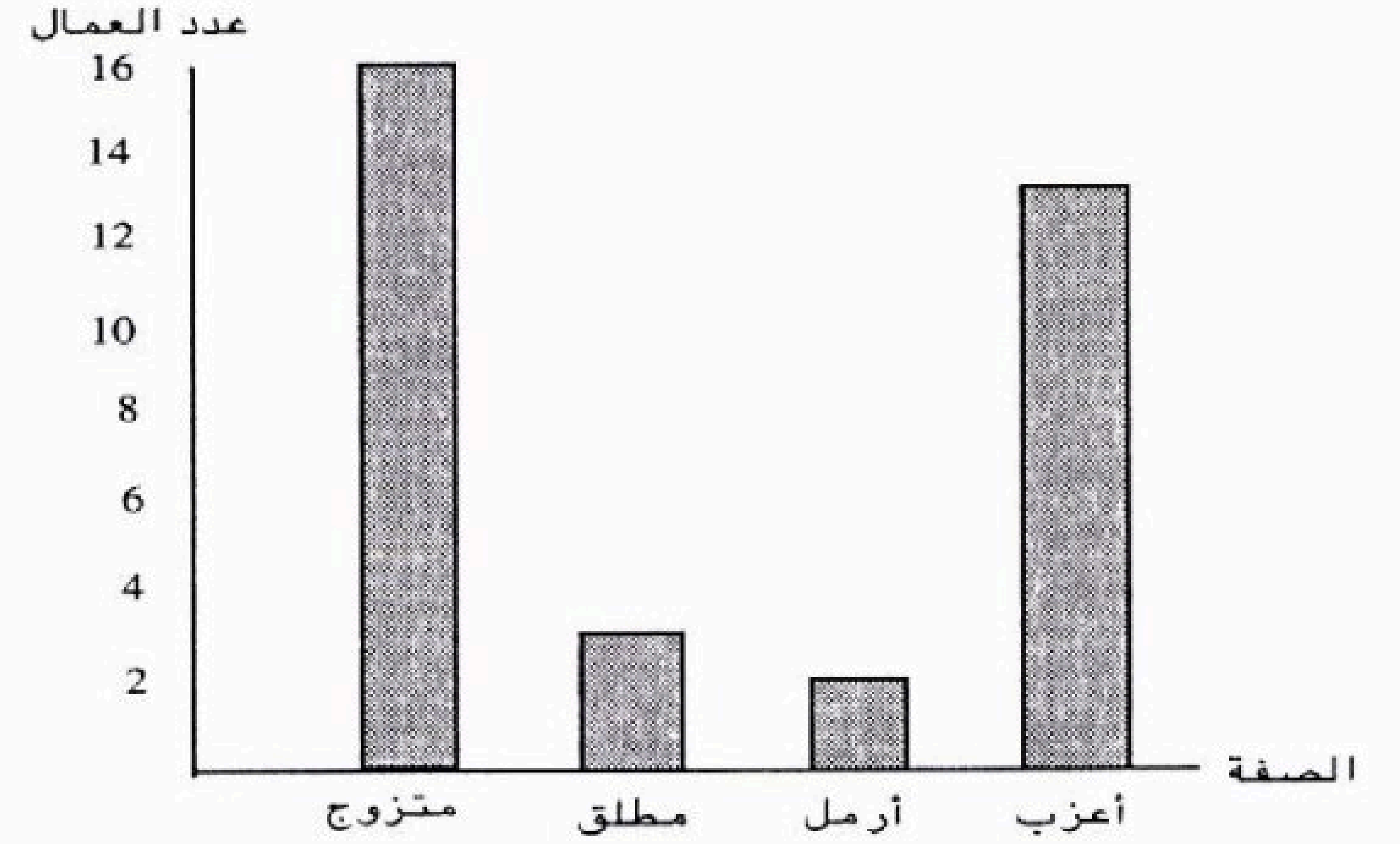
السنة	1395/96	1396/97	1397/98	1398/99	13999/400	1400/401
عدد مدارس الذكور	177	209	273	322	343	375
عدد مدارس الإناث	35	48	58	85	113	138

الحل

ا) نكتب جدول الحالة الإجتماعية من مثال (١,٢) السابق كما يلي:

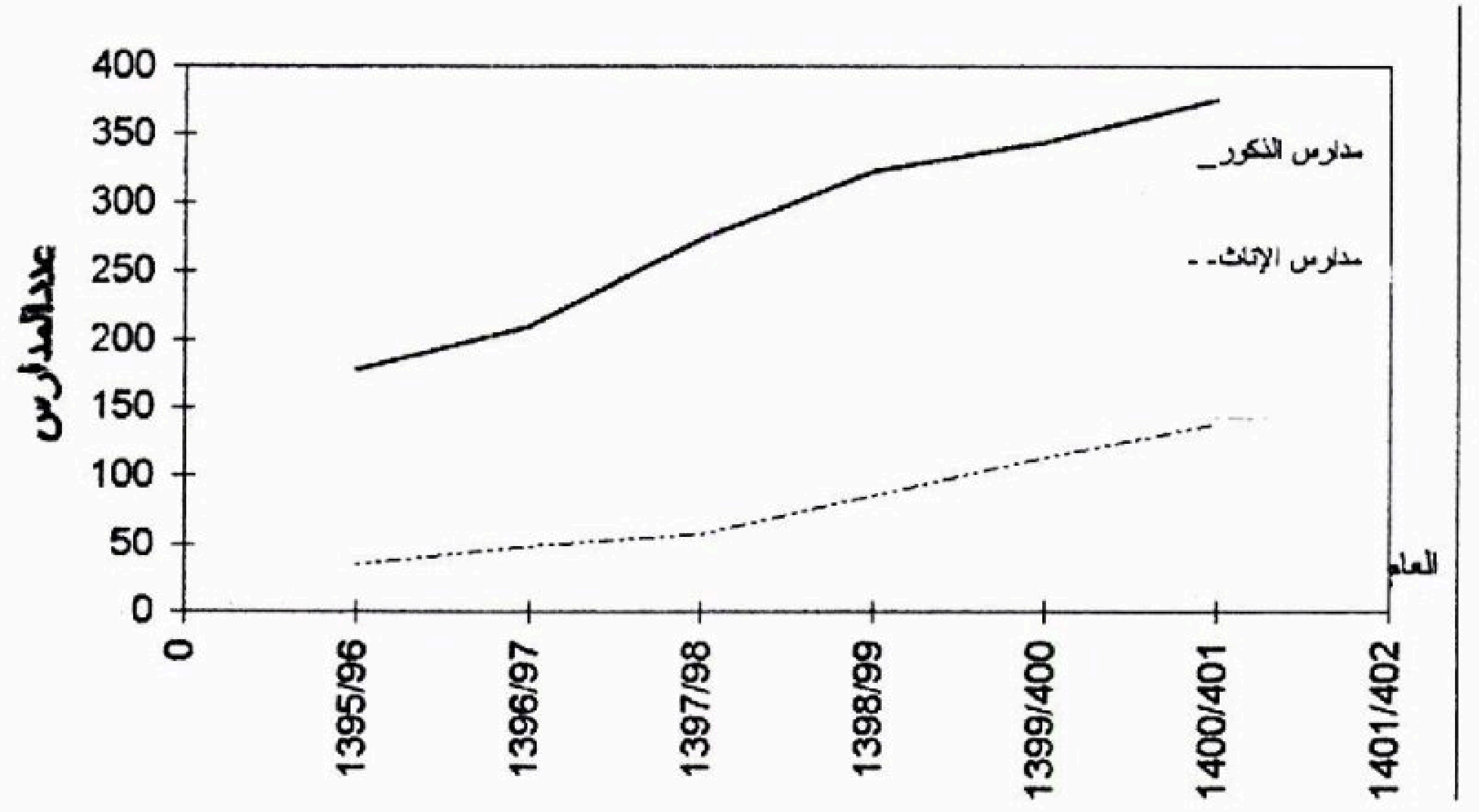
الصفة الإجتماعي	متزوج	مطلق	أرمل	أعزب	المجموع
عدد العمال	16	3	2	9	30

ونرسم محورين متعامدين الأفقي يمثل الصفة والرأسي يمثل التكرار (عدد العمال) وفي الأعمدة يختار سمك مناسب فوق الصفة ارتفاعه يساوي تكرار هذه الصفة كما هو موضح في شكل (١,٢) كما يلي:



شكل (١,٢) التمثيل بالأعمدة البيانية

ب) ويمثل الجدول السابق بالخطين البيانيين الآتيين: كما في شكل (٢,٢) التالي:



شكل (٢,٢) التمثيل بالخط البياني

المدرج التكراري (histogram)

يرسم المدرج التكراري على محورين متعامدين الأفقي يمثل حدود الفئات الحقيقية والرأسي يمثل التكرار ونرسم مستطيلات متلاصقة قاعدتها طول الفئة وارتفاعاتها تكرارات هذه الفئات كما نوضح ذلك في شكل (٢,٣) كما يلي .

المضلع التكراري (Polygon)

مثل المدرج التكراري على محورين متعامدين وبدلا من رسم المستطيلات نرسم نقاط أو دوائر صغيرة أعلى مراكز الفئات ارتفاعاتها تمثل تكرارات هذه الفئات ثم نصل هذه النقاط أو الدوائر بمستقيمات تصل بين كل نقطتين متتاليتين بالمسطرة لنحصل على المضلع التكراري ولغلقه نتخيل فئة وهمية صغرى تكرارها

صفرا نصل مركزها بارتفاع الفئة التالية لها . وكذلك فئة وهمية كبرى ارتفاع تكرارها صفرا نصل مركزها بارتفاع مركز الفئة السابقة لها كما هو موضح في شكل (٤ , ٢).

المنحنى التكراري (frequency curve)

مثل المضلع التكراري وبدلا من توصيل النقاط بالمسطرة يتم ذلك باليد لنحصل على منحنى ممهد خالي من التكسرات أو الزوايا حتى لو اضطررنا بعدم المرور ببعض النقاط كما نوضح ذلك في شكل (٥ , ٢) كما يلي :

المضلع المتجمع الصاعد أو الهابط (increasing or decreasing plogans)

يرسم المضلع المتجمع الصاعد من جدول المتجمع الصاعد والمضلع المتجمع الهابط من الجدول المتجمع الهابط وذلك على محورين متعامدين الأفقي يمثل حدود الفئات الحقيقية والرأسي يمثل التكرار المتجمع وهو موضح في شكل (٦ , ٢) التالي.

مثال (٥ , ٢)

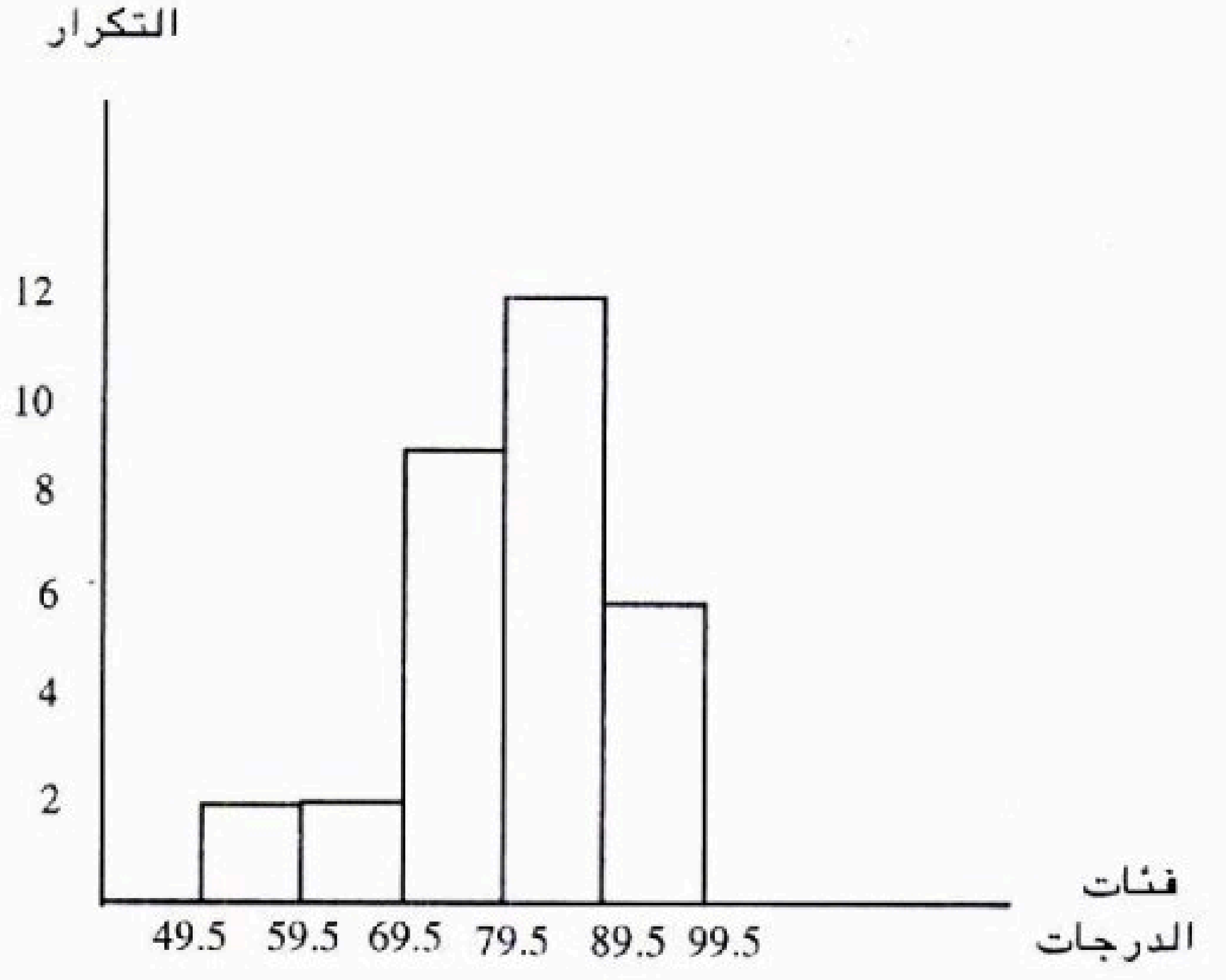
ارسم المدرج والمضلع والمنحنى التكراري لدرجات الطلاب من مثال (٣ , ٢) السابق ثم ارسم المضلع المتجمع الصاعد والهابط .

الحل

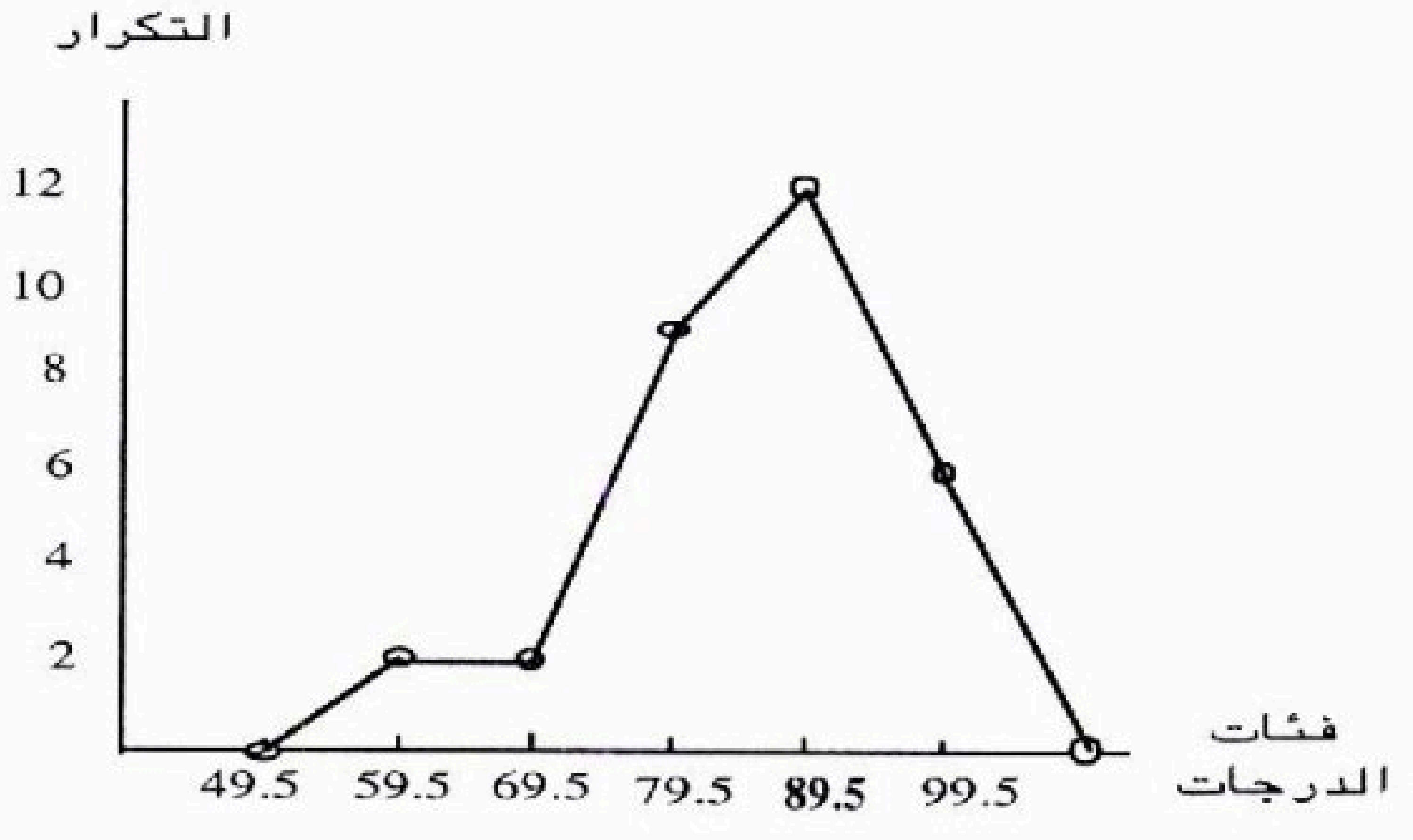
نكتب جدول (٦ , ٢) لتوزيع الدرجات في مثال (٣ , ٢) السابق كما يلي :

فئات الدرجات	50-59	60-69	70-79	80-89	90-99	المجموع
التكرار (عدد الطلاب)	2	2	9	11	6	30

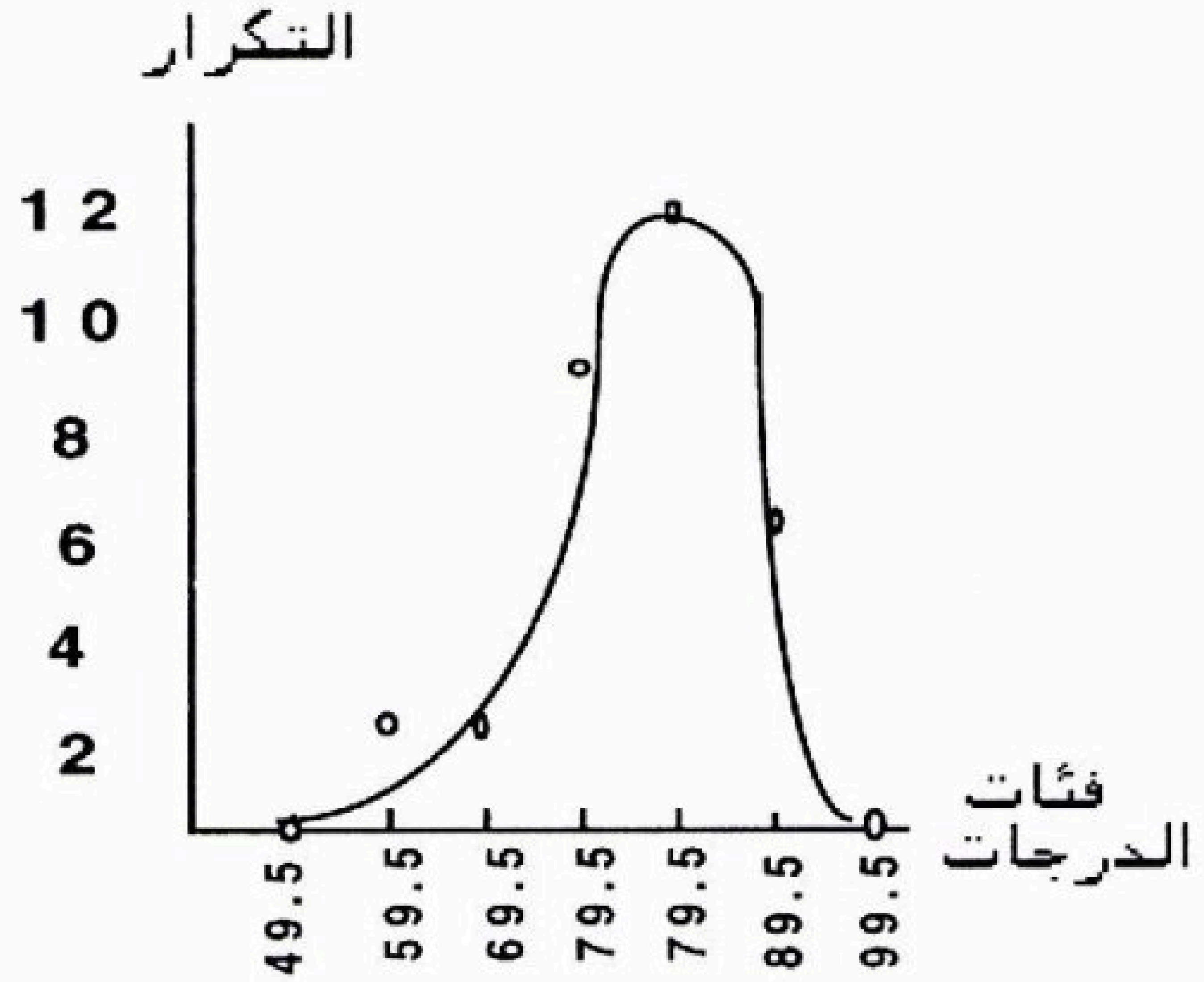
الأشكال التالية تمثل المدرج والمضلع والمنحنى التكراري للتوزيع التكراري لدرجات 30 طالبا كما يلي :



شكل (٣ , ٢) يمثل المدرج التكراري



شكل (٤ , ٢) يمثل المضلع التكراري



شكل (٥ , ٢) يمثل المنحنى التكراري

ثم نرسم المضلع التكراري المتجمع الصاعد شكل (٦ , ٢) والمضلع التكراري المتجمع الهابط شكل (٧ , ٢) كما يلي بعد تكوين الجدولين المتجمعين الصاعد والهابط رقمي (١٠ , ٢) و (١١ , ٢) كما يلي:

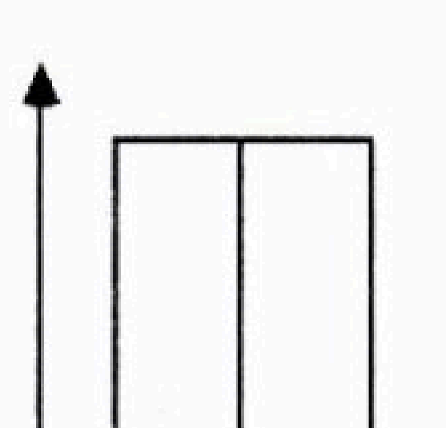
جدول (١١ , ٢) يبين التكرار المتجمع الهابط جدول (١٠ , ٢) يبين التكرار المتجمع الصاعد

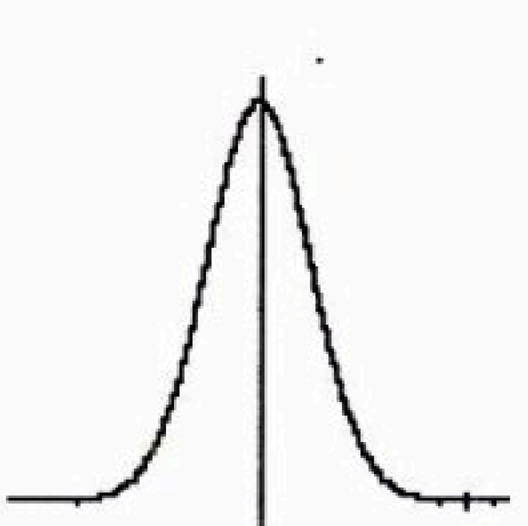
تكرار متجمع هابط	حدود الفئات الدنيا
30	أكبر من .49
28	» » 59.5
26	» » 69.5
17	» » 79.5
6	» » 89.5
0	» » 99.5

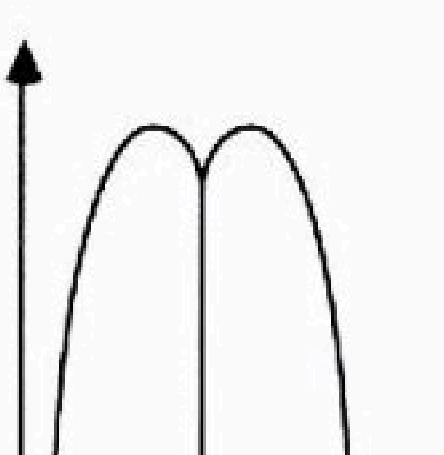
تكرار متجمع صاعد	حدود الفئات الدنيا
0	أل من 49.5
2	» » 59.5
4	» » 69.5
13	» » 79.5
24	» » 89.5
30	» » 99.5

شكل (٢,٧) يمثل المضلع المتجمع الهابط شكل (٢,٦) يمثل المضلع المتجمع الصاعد

يوجد في الحياة العملية كثير من المنحنيات التكرارية غير المتماثلة وقليل من المنحنيات المتماثلة التي لها محور تناظر نستعرض بعض المنحنيات المتماثلة في الأشكال (٢,٨) ، (٢,٩) ، (٢,١٠) و (٢,١١) كما يلي :

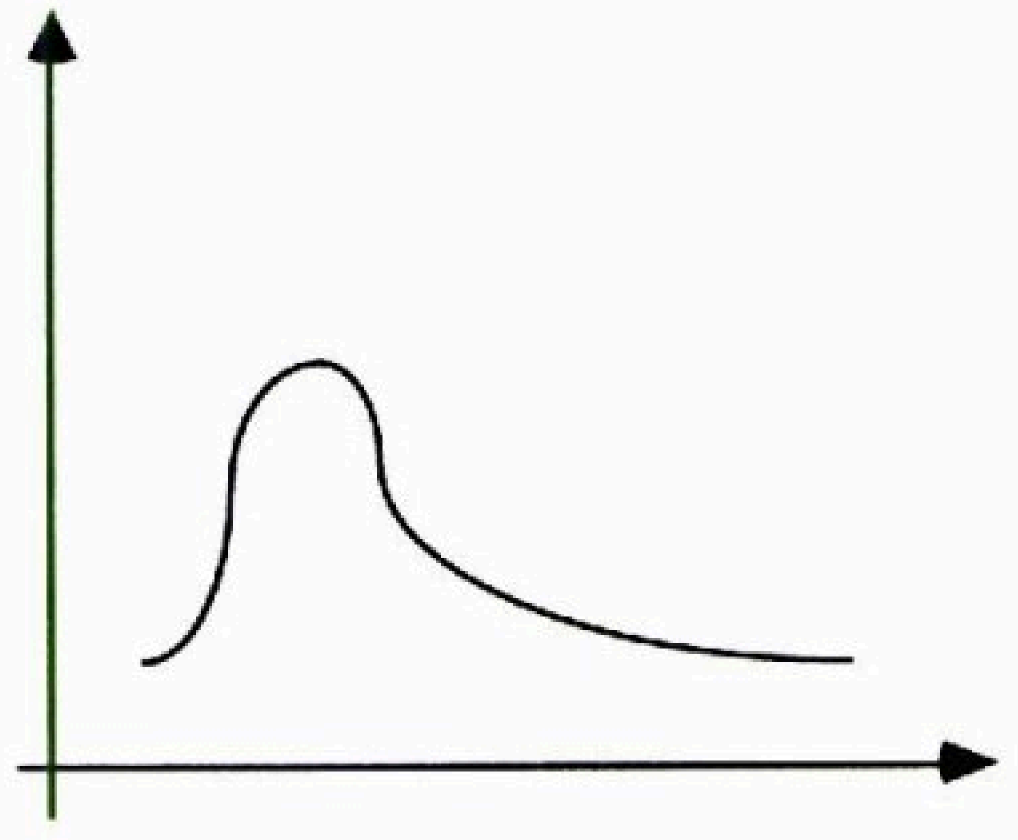




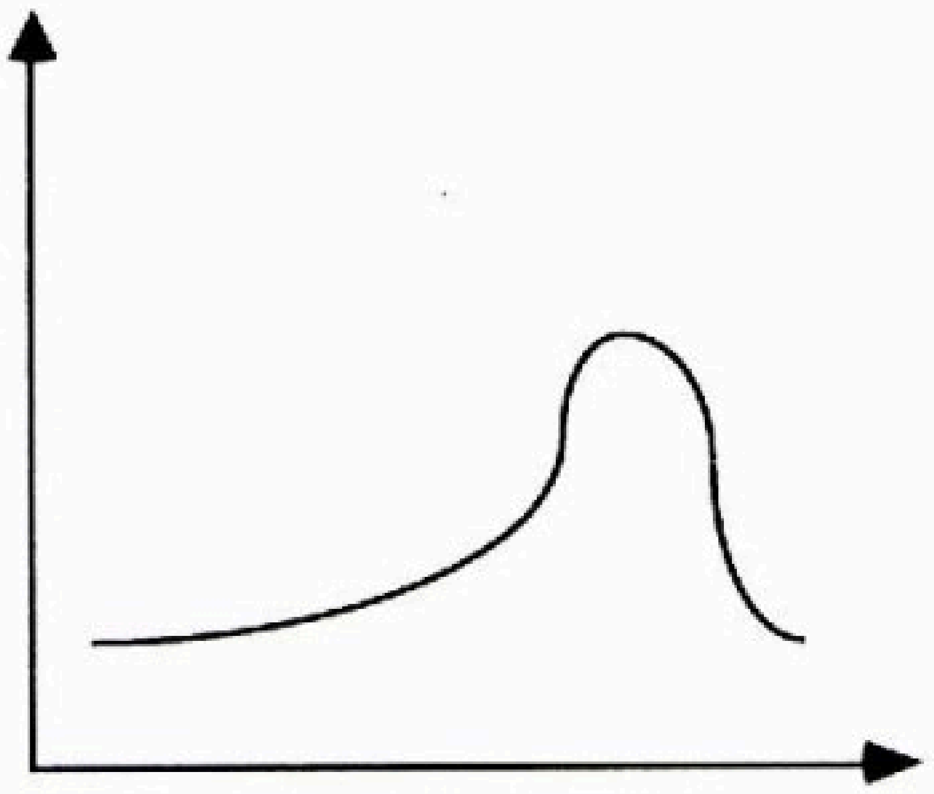


شكل (٢,١١) شكل (٢,١٠) شكل (٢,٩) شكل (٢,٨)

ونستعرض بعض المنحنيات غير المتماثلةالتي ليس لها محور تناظر في الشكلين (٢,١٢) و (٢,١٣) التاليين :



شكل (١٢ , ٢) ملتو نحو اليمين
موجب الالتواء



شكل (١٣ , ٢) ملتو نحو اليسار
سالب الالتواء

(٤ , ٢) مقاييس النزعة المركزية (المتوسطات)
Averages and Measures of Control Tendency

سبق أن استعرضنا طرق تلخيص البيانات جدوليا ، وقمنا بتمثيل البيانات بيانيا بهدف الحصول على بعض الخصائص للمجتمع الإحصائي محل الدراسة من خلال شكل المنحنيات التكرارية بصورة عامة . دعت الحاجة إلى دراسة مقاييس عددية محددة للمقارنة بين أشكال التوزيعات المختلفة ، ومن هذه المقاييس مقاييس النزعة المركزية (أو المتوسطات) وهي عبارة عن قيم مثلى تقترب منها معظم البيانات الإحصائية أو تتركز حولها . وبعض المتوسطات الأكثر شيوعا والتي سوف نتناولها بالشرح والأمثلة هي: الوسط الحسابي، الوسط المرجح، المنوال، الوسيط والمئينات مع ذكر أهم مميزات وعيوب كل منها .

(١ , ٤ , ٢) الوسط الحسابي (Arithmetic mean)

يعتبر الوسط الحسابي من أهم مقاييس النزعة المركزية والأكثر استخداما إذا كان لدينا متغيرا X له عدد من المشاهدات $x_1, x_2, \dots, x_n$. فيعرف متوسط

هذه المشاهدات ويرمز له بالرمـز $\overline{x}$ بأنه يساوي مجموع هذه المشاهدات عـلـى عددها أي أن :

$$\overline{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$

أي أن :

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad , \quad i = 1, 2, ..., n \qquad (٢ , ١)$$

مثال (٥ , ٢)

إذا كانت أوزان مجموعة من خمسة طلاب بالمرحلة الابتدائية هي :

42, 40, 44, 48, 46

احسب الوسط الحسابي لأوزان هؤلاء الطلاب .

الحل

$\overline{x} = (42 + 40 + 44 + 48 + 46) / 5 = 44$ كجم

ويعرف الوسط الحسابي للبيانات المبوبة (data grouped) كما يلي :

إذا كان لدينا عدد k من الفئات لها مراكز $x_1, x_2, ..., x_k$ ولها تكرارات $f_1, f_2, ..., f_k$ على الترتيب فيكون الوسط الحسابي $\overline{x}$ كما يلي :

$$\overline{x} = \frac{f_1 x_1 + f_2 x_2 + ... + f_k x_k}{f_1 + f_2 + ... + f_k}$$

أي أن

$$\overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{n} \qquad (٢ , ٢)$$

حيث إن $n = f_1 + f_2 + ... + f_k$

مثال (٢,٦)

احسب الوسط الحسابي $\overline{x}$ لدرجات الطلاب في مثال (٢,٣) السابق .

الحل

لسهولة الحل نضع الجدول التكراري لتوزيع درجات الطلاب كما يلي :

فئات الدرجات	مراكز الفئات x	التكرار f	f x
50 - 59	54.5	2	109
60 - 69	64.5	2	129
70 - 79	74.5	9	760.5
80 - 89	84.5	11	929.5
90 - 99	94.5	6	567.0
المجموع Σ		30	2405

$$\overline{x} = \Sigma f x / n$$

$= 2405/30 = 80.17$ درجة

بعض خصائص الوسط الحسابي

١) مجموع انحرافات القيم عن وسطها الحسابي يساوي صـفـرا . أي أن $\Sigma(x - \overline{x}) = 0$ أي أن :

$$\Sigma(x - \overline{x}) = \Sigma x - n\overline{x} = \Sigma x - n\left(\frac{\Sigma x}{n}\right) = 0 .$$

٢) إذا طرحنا أو أضفنا مقدار ثابـت a فيكون الوسط الحسابي $\overline{x}$ للقيـم الأصلية يكون مساويا للوسط الجديد $\overline{d}$ يضاف اليه أو مطروحا منه المقدار الثابت a أي أن $\overline{x} = \overline{d} \pm a$.

أولا في حالة طرح المقدار الثابت a أي أن

$$d_i = x_i - a \quad , \quad i = 1, 2, ..., n$$

$$\Leftrightarrow \quad \Sigma d_i = \Sigma x_i - n a$$

$$\Leftrightarrow \quad \Sigma d_i / n = \Sigma x_i / n - a$$

$$\Leftrightarrow \quad \bar{d} = \bar{x} - a$$

$$\Leftrightarrow \quad \bar{x} = \bar{d} + a$$

ثانيا في حالة إضافة المقدار الثابت a

نحصل على $\bar{x}$ كما تم في حالة الطرح كما يلي :

$$\bar{x} = \bar{d} - a$$

٣) إذا ضربنا أو قسمنا مقدار ثابت a لقيم المتغير x فإننا نحصل على قيم متغير جديد y حيث $y_i = a x_i \; , \; i = 1 , 2, ... n$ ويحسب المتوسط الجديد $\bar{y}$ من العلاقة $\bar{y} = a\,\bar{x}$ وذلك في حالة الضرب ، أما حالة القسمة من العلاقة $\bar{y} = \frac{\bar{x}}{a}$.

مثال (٢،٧) :

احسب $\bar{x}$ في مثال (٢،٥) وذلك باستخدام الخاصية(١) و (٢) عند طرح مقدار ثابت $a = 44, \; a=40$ وكذلك في حالة الخاصية ٣ عندما $a = \frac{1}{2}$

الحل

بأخذ $a = 44$ حيث قيم x هي 46، 48، 44، 40، 42 نجد أن الخاصية (١) محققة كما يلي :

$$\Sigma(x_i - 44) = (42\text{-}44) + (40\text{-}44) + (48\text{-}44) + (46\text{-}44)$$
$$= - 2 - 4 + 4 + 2 = 0$$

وبأخذ $a = 40$ نجد أن الخاصية (٢) محققة كما يلي :

$$\bar{x} = \bar{d} + a$$
$$= \frac{2+0+4+8+6}{5} + 40$$
$$= 4 + 40 = 44$$

وبأخذ $a = \frac{1}{2}$ نجد أن الخاصية (٣) لحساب المتوسط الجديد $\bar{y}$ هو كما يلي:

$$\bar{y} = a\,\bar{x} = \frac{1}{2}(44) = 22$$

(١ , ١ , ٤ , ٢) مميزات الوسط الحسابي

مقياس سهل حسابه وتأخذ جميع القيم أثناء حسابه ويسهل استخدامه في التحليل الإحصائي .

(٢ , ١ , ٤ , ٢) عيوب الوسط الحسابي

- يتأثر بالقيم الشاذة نحو الصغر أو الكبر وخاصة عندما يكون عدد البيانات صغير نسبيا ويقل التأثر كلما زاد عدد البيانات .

- يصعب حسابه في حالة الجداول التكرارية المفتوحة وذلك لعدم تحديد مركز الفئة المفتوحة ونوضح ذلك بفتح جدول درجات الطلاب (٦ , ٢) السابق من أسفل كما هو موضح فيما يلي :

فئات الدرجات	59 فأقل	60 - 69	70 - 79	80 - 89	90 - 99
عدد الطلاب	2	3	9	11	6

ويصعب تحديد مركز الفئة المفتوحة (59 فأقل) لعدم معرفة الحد الأدنى لهذه الفئة .

- يصعب حسابه في حالة البيانات الوصفية .

(٢,٤,٢) الوسط المرجح أو الموزون (Weighted mean)

أحيانا بعض القيم $x_1, x_2, ..., x_k$ تكون مقرونة بأهمية أو أوزان $w_1, w_2, ..., w_k$ على الترتيب فيكون الوسط المرجح $\overline{x}_w$ يعطى بالعلاقة التالية :

$$\overline{x}_w = \frac{w_1 x_1 + w_2 x_2 + ... + w_k x_k}{w_1 + w_2 + ... + w_k} = \frac{\Sigma w x}{\Sigma w} \qquad (٢,٣)$$

مثال (٢,٨)

احسب الوسط المرجح لدرجات شعبتين لمقرر ١٠٣ إحص حيث كان لدينا المعلومات التالية :

الشعبة الأولى : درجة $\overline{x}_1 = 80$ وعدد طلابها $n_1 = 40$

الشعبة الثانية : درجة $\overline{x}_2 = 70$ وعدد طلابها $n_2 = 60$

الحل

نختار الأوزان كما يلي :

$$w_1 = n_1 = 40$$
$$w_2 = n_2 = 60$$

أي أن الوسط المرجح لدرجات الشعبتين هو

$$\overline{x}_w = \frac{\overline{x}_1 w_1 + \overline{x}_2 w_2}{w_1 + w_2} = \frac{80(40) + 70(60)}{40 + 60}$$

أي أن

$$\overline{x}_w = 7400 / 100 = 74 \quad \text{درجة}$$

ملاحظة

يمكن اعتبار الوسط الحسابي في الجداول التكرارية وسط مرجح باعتبار

تكرارات الفئات كأوزان لهذه الفئات أي أن :

$$w_i = f_i \,, \quad i = 1, 2, \ldots, k$$

أي أن

$$\bar{x} = \bar{x}_w = \frac{\Sigma w \bar{x}}{\Sigma w} = \frac{\Sigma f x_i}{\Sigma f}$$

(٣ , ٤ , ٢) الوسيط (Median)

يعرف الوسيط بالقيمة التي يسبقها 50% من المشاهدات بعد ترتيبها . فإذا كان عدد المشاهدات فردى فالوسيط هو المشاهدة التي في المنتصف وإذا كان عدد المشاهدات زوجي فإن الوسيط هو متوسط المشاهدتين اللتين في المنتصف .

مثال (٩ , ٢)

احسب الوسيط لأوزان الطلاب في مثال (٥ , ٢) السابق .

الحل

نرتب أوزان الطلاب تصاعديا كما يلي :

40, 42, 44, 46, 48

رتبة الوسيط هي= $\frac{n+1}{2} = \frac{5+1}{2} = 3$

الوسيط $\tilde{x}$ هو قيمة القراءة الثالثة بعد الترتيب :

$\tilde{x} = 44$ كجم

مثال (١٠ , ٢)

احسب الوسيط لأوزان 6 طلاب أوزانهم هي :

42, 40, 44, 45, 48, 46

الحل

ترتب البيانات تصاعديا كما يلي :

40, 42, 44, 45, 46, 48

الوسيط $\tilde{x}$ هو الوسط الحسابي للقيمتين اللتين رتبتهما هما $\frac{n}{2}$ ، $\frac{n+2}{2}$ أي

$\frac{6}{2}$ ، $\frac{6+2}{2}$ أي الرتبتين 4, 3 .

الوسيط $\tilde{x}$ هو الوسط الحسابي للقيمتين اللتين رتبتهما الثالثة والرابعه أي أن

$$\tilde{x} = \frac{44+45}{2} = 44.5 \quad \text{كجم}$$

(١ , ٣ , ٤ , ٢) الوسيط في حالة الجداول التكرارية (البيانات المبوبة).

يحسب الوسيط في هذه الحالة من الجدول التكراري المتجمع الصاعد حيث تحدد الفئة التي تشتمل على عدد يمثل نصف التكرارات $\left(\frac{n}{2}\right)$ وتسمى بالفئة الوسيطية ويحسب الوسيط $\tilde{x}$ بالعلاقة التالية :

$$\tilde{x} = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} . L \qquad (٢ , ٤)$$

حيث A تمثل بداية الفئة الوسيطية .

$\frac{n}{2}$ تمثل نصف التكرارات سواء كانت n فردية أم زوجية .

f_1 تمثل التكرار المتجمع الصاعد الذي يسبق الفئة الوسيطية .

f_2 تمثل التكرار اللاحق المتجمع الصاعد الذي يلي الفئة الوسيطية

L يمثل طول الفئة الوسيطية .

مثال (٢,١١)

احسب الوسيط $\tilde{x}$ لدرجات الطلاب من الجدول التكراري في مثال (٢,٣) السابق . ثم احسب ($\tilde{x}$) بيانيا .

الحل

نكون الجدول التكراري المتجمع الصاعد ونضع خط أفقي عند منتصف التكرارات أي أن $15 = \frac{30}{2} = \frac{n}{2}$ كما يلي :

التكرار المتجمع	حدود الفئات الدنيا
0	أقل من 49.5
2	59.5 » »
4	69.5 » »
13 f_1	A 79.5 » »
24 f_2	89.5 » »
30	99.5 » »

أولا: الوسيط $\tilde{x}$ حسابيا

من الجدول نجد أن

$$A = 79.5 \ , \ f_1 = 13 \ , \ f_2 = 24 \ , \ L = 10$$

ومن العلاقة (٢,٤) الوسيط $\tilde{x}$ هو

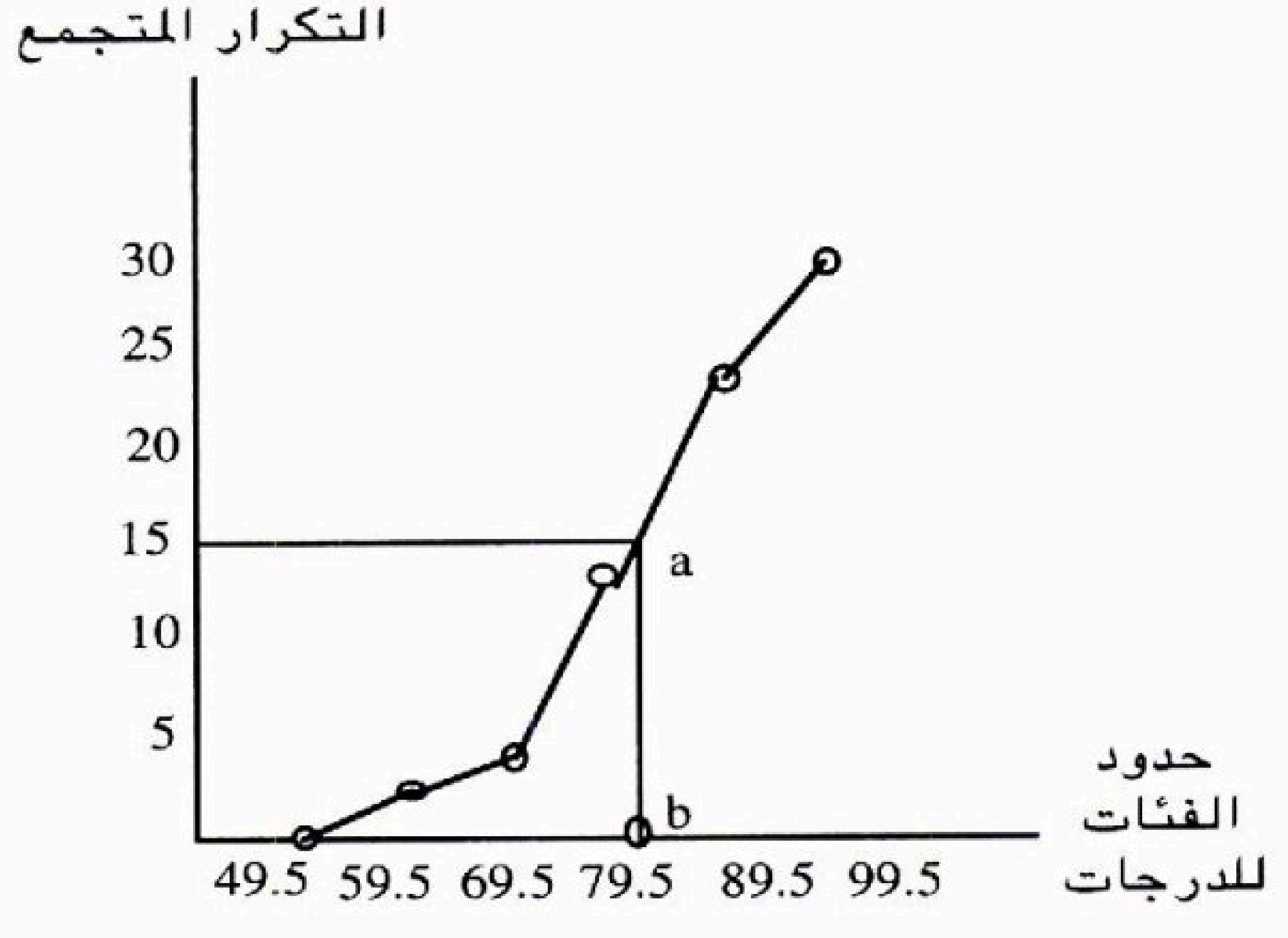
$$\tilde{x} = 79.5 + \frac{15 - 13}{24 - 13} . 10$$

$$= 79.5 + 1.8$$

$$= 81.3 \quad \text{درجة}$$

ثانيا: الوسيط بيانيا

نرسم المضلع المتجمع الصاعد وعلى محور التكرارات نحدد تكرار الوسيط $15 = \frac{30}{2}$. ثم نرسم خط أفقي يقطع المضلع المتجمع الصاعد في نقطة a ثم نسقط من a عمود منقط رأسي يلتقي بمحور حدود الفئات في نقطة b تكون قيمتها على تدريج محور الفئات هي قيمة الوسيط ونوضح ذلك في شكل (٢,١٤) التالي:



شكل (٢,١٤)

قيمة الوسيط $\tilde{x}$ عند b هو

$$\tilde{x} = 79.5 + 2 = 81.5 \quad \text{درجة}$$

ملاحظة :

يمكن حساب الوسيط حسابيا من الجدول المتجمع الهابط وكذلك بيانيا من

المضلع المتجمع الهابط ، كما أن مسقط نقطة تلاقي المضلعين الصاعد والهابط تمثل قيمة الوسيط $\tilde{x}$ على محور فئات الدرجات .

(٢ , ٣ , ٤ , ٢) مميزات الوسيط

١ - قياس سهل حسابه لا يتأثر بالقيم المتطرفة نحو الكبر أو الصغر.

٢ - يمكن حسابه في حالة الجداول التكرارية المفتوحة من أسفل أو أعلى .

٣ - يمكن حسابه للبيانات الوصفية التي لها صفة التراتيب مثل تقديرات الطلاب (ضعيف - مقبول - جيد - جيدجدا - ممتاز) .

(٣ , ٣ , ٤ , ٢) عيوب الوسيط

١ - لا يأخذ جميع القيم في الأعتبار عند حسابه .

٢ - لا يسهل التعامل معه في االتحليل الإحصائي .

(٤ , ٤ , ٢) المنوال (Mode)

يعرف المنوال بأنه هو القيمة الأكثر شيوعا (تكرارا) في مجموعة القراءات محل الدراسة وقد يكون للقراءات أو المشاهدات منوالا واحدا أو أكثر من منوال أو لا يوجد لها أية منوال ونوضح ذلك بالأمثلة التالية :

مثال (١٢ , ٢)

احسب المنوال لأوزان مجموعة من طلاب الإبتدائي كانت أوزانهم هي :

42, 45, 43, 44, 43, 44, 46, 44

الحل

نختار القراءة التي لها أكبر تكرار فهي تمثل المنوال $\hat{x}$ أي أن

$\hat{x} = 44$ كجم

حيث 44 تكررت 3 مرات وهو أكبر تكرار للبيانات .

مثال (١٣ , ٢)

احسب المنوال لأوزان مجموعة من طلاب الإبتدائي حيث أوزانهم :

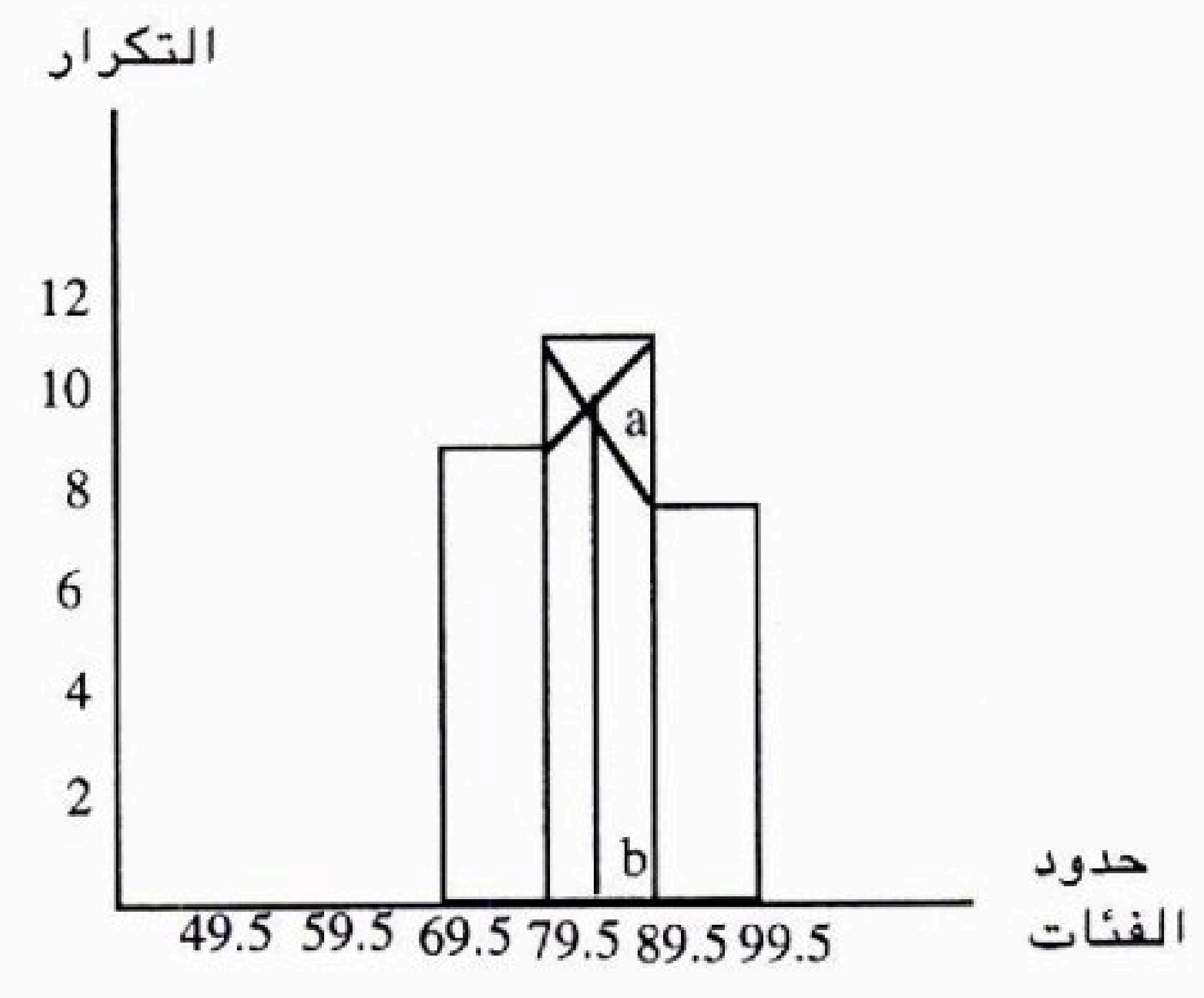
42, 45, 43, 44, 43, 44, 46, 43, 44

الحل

نختار القراءتين اللتان لهما أكبر تكرار وهما 44 و 43 كجم أي يوجد لهذه البيانات أكثر من منوال وهما 44 و 43 لأن كل منهما تكرر 3 مرات وهو يمثل أكبر تكرار للبيانات .

ملاحظة

يمكن حساب المنوال في مثال (١٥ , ٢) بطريقة الرافعة ونوضح ذلك بيانيا بأن نرسم المدرج التكراري أو جزء منه أي ثلاث فئات فقط الفئة المنوالية والسابقة والتالية لها . نصل نهاية قمة مستطيل الفئة السابقة بنهاية قمة المستطيل للفئة المنوالية، ونصل بداية قمة مستطيل الفئة اللاحقة ببداية قمة المستطيل للفئة المنوالية. نحصل على نقطة تقاطع (a) مثلا ، يكون مسقط (a) على محور الفئات وليكن (b) هو قيمة المنوال ونوضح ذلك في الشكل (١٥ , ٢) كما يلي :



شكل (١٥ , ٢)

من الرسم قيمة المنوال (b) = 79.5 + 2 = 81.5 درجة تقريبا .

(١ , ٤ , ٤ , ٢) مميزات المنوال

١ - مقياس سهل حسابه ولا يتأثر بالقيم المتطرفة نحو الصغر أو الكبر .

٢ - يسهل حسابه في حالة الجداول التكرارية المفتوحة من الطرفين.

٣ - يسهل حسابه في حالة البيانات الوصفية عموما سواء كانت لها صفة التراتيب مثل تقديرات الطلاب (ضعيف ، مقبول ، جيد ، جيدجدا ، ممتاز) أو ليست لها تراتيب مثل الحالة الاجتماعية (أعزب - متزوج - أرمل - مطلق).

(٢ , ٤ , ٤ , ٢) عيوب المنوال

١ - مقياس غير دقيق لأنه لا يأخذ جميع القيم أثناء حسابه .

٢ - قد يكون للبيانات أكثر من منوال أو عديمة المنوال .

٣ - قد لا يأتي المنوال في مركز البيانات بل قد يأتي في أحد الأطراف مما يعتبره بعض الإحصائيين أنه ليس مقياس للنزعة المركزية .

مثال (١٤ , ٢)

احسب المنوال لأوزان مجموعة طلاب الابتدائي في مثال (٥ , ٢) حيث كانت الأوزان هي :

42, 40, 44, 48, 46

الحل

لا توجد لهذه القراءات أية منوال لأنه لا يوجد قراءة لها تكرار أكبر من القراءات الباقية في هذه الأوزان .

(٣ , ٤ , ٤ , ٢) المنوال في حالة الجداول التكرارية (البيانات المبوبة)

يحسب المنوال في هذه الحالة بطريقة مقربة ومبسطة سواء كان حسابيا أو

بيانيا بقيمة مركز الفئة المنوالية . حيث تعرف الفئة المنوالية بأنها الفئة التي يناظرها أكبر تكرار .

مثال (٢,١٥)

احسب المنوال لدرجات الطلاب في مثال (٢,٣) السابق .

الحل

نعيد كتابة الجدول التكراري في مثال (٢,٣) السابق كما يلي :

فئات الدرجات	50 - 59	60 - 69	70 - 79	80 - 89	90 - 99
عدد الطلاب	2	2	9	11	6

الفئة المنوالية هي (89 - 80) حيث تكرارها هو أكبر تكرار ويساوي 11 طالبا ويكون مركزها تقريبا مساويا لقيمة المنوال $\hat{x}$ لهذه القراءات أي أن :

$$\hat{x} = \frac{80+89}{2} = 84.5 \quad \text{درجة}$$

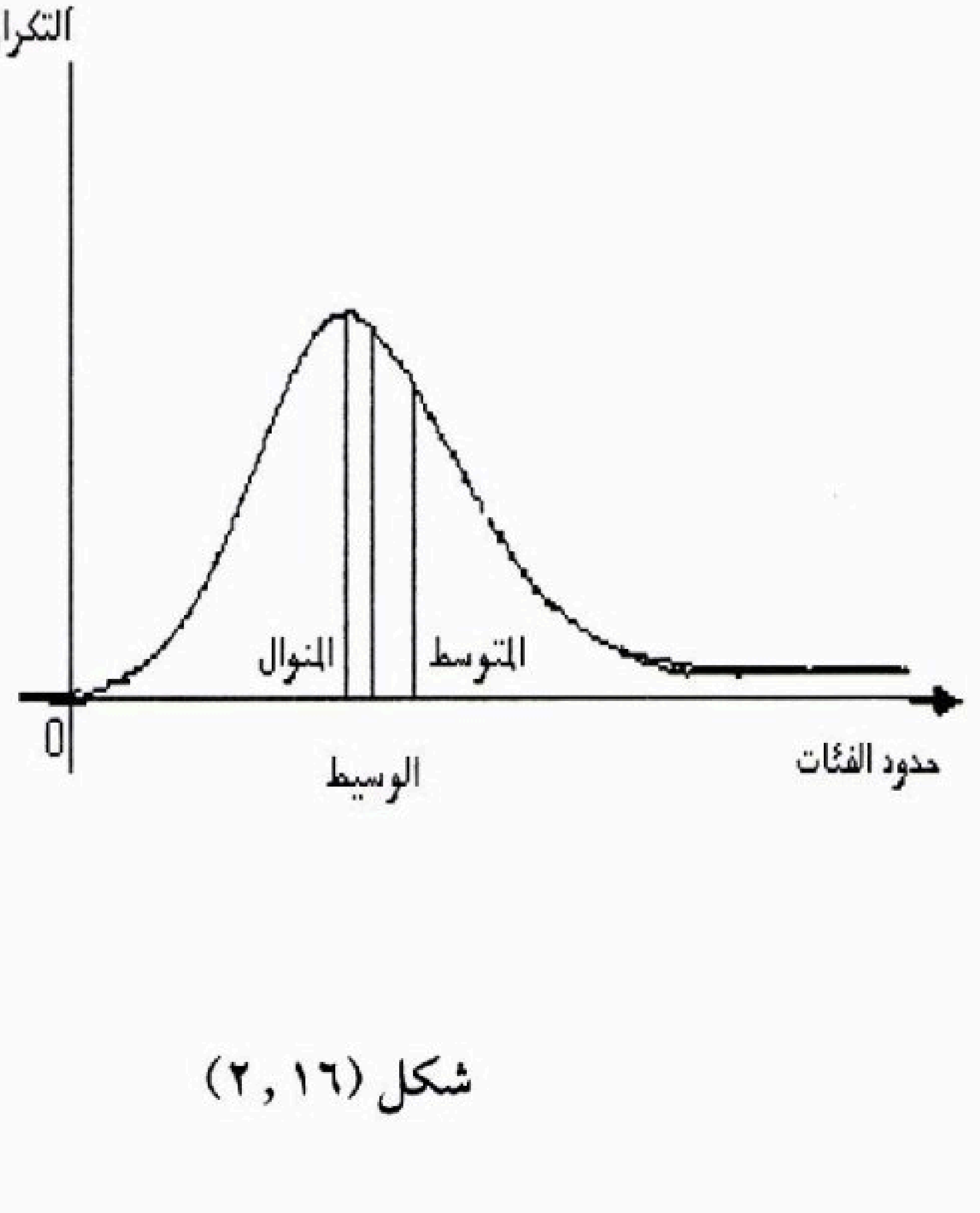
ملاحظات

(١) نلاحظ في الأمثلة السابقة (٢,٦) ، (٢,١١) و (٢,١٥) الخاصة بالوسط الحسابي ($\bar{x}$) الوسيط ($\tilde{x}$) و المنوال ($\hat{x}$) حيث كانت هي :

$\hat{x} = 84.5$ درجة ، $\tilde{x} = 81.3$ درجة ، $\bar{x} = 80.17$ درجة .

يتضح من هذه النتائج أن مقاييس النزعة المركزية (المتوسطات) تعطي إنطباعا واحدا وأن متوسط الدرجات لهؤلاء الطلاب هو حول الثمانين درجة. أي أن متوسط درجاتهم هو جيد جدا لمقاييس النزعة المركزية السابقة.

(٢) في حالة التوزيعات التكرارية المتماثلة الوحيدة المنوال تتطابق قيم المقاييس الثلاثة السابقة وتعطي قيمة واحدة . أما في حالة الالتواء نحو اليمين نجد أن قيمة الوسيط تقع بين المنوال والوسط الحسابي كما هو مبين بالشكل (٢, ١٦) كما يلي:



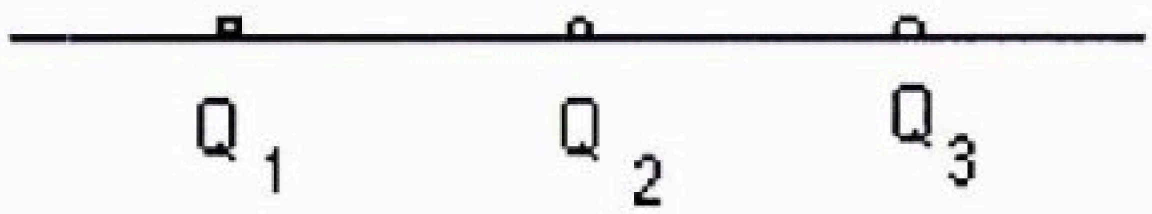
شكل (٢, ١٦)

(٢, ٥) الربيعات والعشيرات والمئينات

(Quantiles, Decimals, and Percentile)

إذا رتبت مجموعة من البيانات تصاعديا أو تنازليا فإن القيمة التي تكون في منتصف البيانات تسمى الوسيط كما سبق دراسته في بند (٢, ٤, ٣, ١) بتعميم فكرة تقسيم البيانات بعد ترتيبها إلى أقسام متساوية تسمى نقط التقسيم الربيعات Q_1, Q_2, Q_3 حيث Q_1 يسمى الربيع الأول وهو الذي يسبقة $\frac{1}{4}$ البيانات،

ويسمى Q_2 الربيع الثاني أو الوسيط وهو الذي يسبقه نصف البيانات ويسمى Q_3 الربيع الثالث وهو الذي يسبقه $\frac{3}{4}$ البيانات، ونوضح ذلك كما يلي:



. كذلك يمكن تقسيم البيانات بعد ترتيبها الى عشرة أقسام متساوية $D_1, D_2, ..., D_9$ حيث يكون D_1 هي القيمة التي يسبقها $\frac{1}{10}$ من البيانات ويسمى العشير الأول وهكذا D_2 هي القيمة التي يسبقها $\frac{2}{10}$ من البيانات وهكذا لباقي العشيرات .

أما بالنسبة للمئينات فإن البيانات تقسم بعد ترتيبها إلى مائة قسم متساوية تسمى $P_1, P_2, ..., P_{99}$ حيث يكون P_1 هي القيمة التي يسبقها $\frac{1}{100}$ من البيانات وهكذا P_2 القيمة التي يسبقها $\frac{2}{100}$ من البيانات وهكذا لباقي المئينات .

مثال (٢ , ١٦)

احسب P_{90} , D_4 , Q_1 في مثال (٢ , ٣) السابق لدرجات الطلاب حسابيا وبيانيا.

الحل

تعمم طريقة حساب الوسيط حيث يمكن حساب المئينات أو العشيرات أو الربيعات حيث الوسيط هو المئين الخمسين P_{50} والعشير الخامس D_5 والربيع الثاني Q_2 . يحسب المئين P_a بالعلاقة (٢ , ٥) التالية ذلك مثل ما تم في حساب الوسيط بالعلاقة (٢ , ٤) السابقة ، ويكون تكرار المئين P_a هو $\frac{a\,n}{100}$ حيث n عدد البيانات .

$$P_a = A + \frac{\frac{a\,n}{100} - f_1}{f_2 - f_1} . L \qquad (٢,٥)$$

حيث A بداية الفئة للمئين P_a .

f_1 التكرار المتجمع الصاعد السابق للفئة المئينية للمئين P_a

f_2 التكرار المتجمع الصاعد اللاحق للفئة المئينية للمئين P_a

L طول الفئة المئينية للمئين P_a .

أي أن الربيع Q_1 يساوي P_{25} أي قيمة $a = 25$ و $n = 30$. في مثالنا تكون النسبة $\frac{a\,n}{100} = \frac{25(30)}{100} = 7.5$ ونضع خط أفقي فى الجدول المتجمع الصاعد التالي بحيث أعلاه قيمة أقل من 7.5 وأسفله قيمة أكبر من 7.5 للتكرارات المتجمعة كما يلي :

التكرار المتجمع الصاعد	حدود الفئات
0	49.5
2	59.5
4	69.5
-------------------- 7.5,12	
13	79.5
25	89.5
--------------------27	
30	99.5

$$Q_1 = P_{25} = 69.5 + \frac{7.5-4}{13-4} . 10 = 73.5 \quad \text{درجة}$$

وبالمثل $P_{40} = D_4$ فإن $a = 40$ والنسبة $\frac{40(30)}{100}$ تساوي 12 فيكون الخط

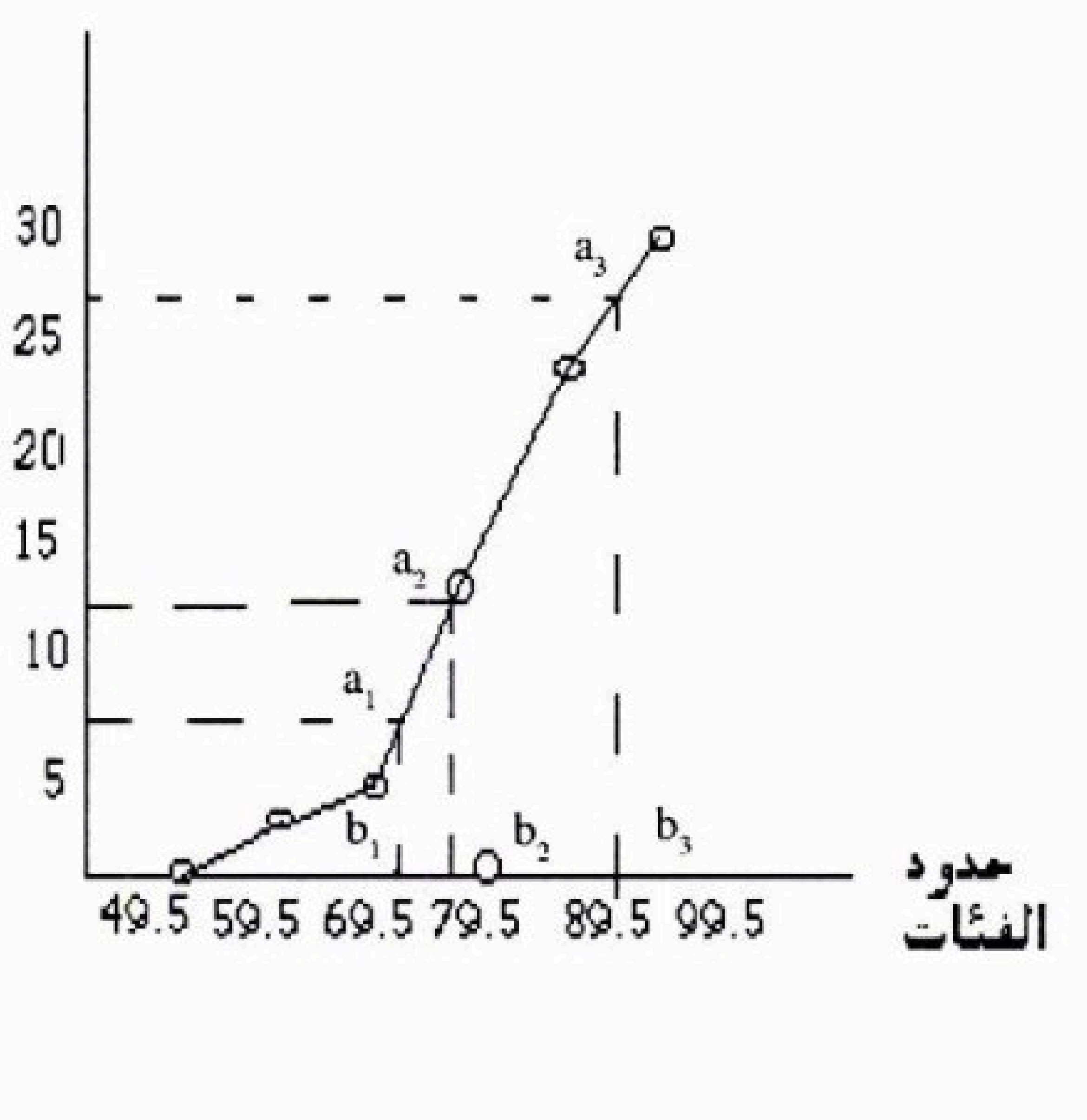
الأفقي نفسه عند 7.5 كما هو موضح بالجدول السابق وتحسب D_4 كما يلي

$$D_4 = P_{40} = 69.5 + \frac{12-4}{13-4} \,.\, 10 = 69.5 + 9 = 78.5$$ درجة

والمئين التسعين P_{90} حيث النسبة $27 = \frac{90(3)}{100}$ والخط الأفقي بين التكرار المتجمع 25 و 30 ويكون

$$P_{90} = 89.5 + \frac{27-25}{30-25} \,.\, 10 = 89.5 + 4 = 93.5$$ درجة

ولحساب Q_1, D_4, P_{90} بيانيا نرسم المضلع المتجمع الصاعد من الجدول المتجمع الصاعد ونرسم خطوط أفقية منقطة للتكرارات 27 , 12 , 7.5 تلتقي مع المضلع المتجمع الصاعد في النقاط a_1, a_2, a_3 تكون مساقطها الرأسية على محور حدود الفئات هي b_1, b_2, b_3 تمثل قيم Q_1, D_4, P_{90} على الترتيب نوضح ذلك في شكل (٢ , ١٧) كما يلي :



شكل (٢ , ١٧)

من الرسم نجد أن :

درجة تقريبا $b_1 = Q_1 = 73$

درجة تقريبا $b_2 = D_4 = 78$

درجة تقريبا $b_3 = P_{90} = 93$

(٦ , ٢) مقاييس التشتت
(Measure of Dispersion)

سبق دراسة تمثيل البيانات الإحصائية بيانيا لمعرفة توزيع الظاهرة محل الدراسة، كما درسنا مقاييس النزعة المركزية لحساب متوسط الظاهرة محل الدراسة، والدراسات السابقة لم تشر إلى معرفة درجة تجانس البيانات الإحصائية التي قد يكون لها نفس المقياس للنزعة المركزية ولكن تختلف البيانات في درجة تجانسها . لذلك دعت الحاجة إلى دراسة مقاييس للتشتت، سندرس منها ثلاثة مقاييس مهمة هي المدى، نصف المدى الربيعي والانحراف المعياري. وهي تصف لنا درجة تقارب أو تباعد البيانات لأهمية ذلك في التحليل الإحصائي واتخاذ القرارات . ولفهم تساوي البيانات في المتوسطات واختلافها في التشتت (عدم التجانس) نورد المثال التالي.

مثال (١٧ , ٢)

احسب متوسط الوزن بالكجم لمجموعتين من طلاب الإبتدائي حيث كانت:

المجموعة الأولى هي: 42, 40, 44, 48, 46

المجموعة الثانية هي: 32, 50, 44, 40, 54

الحل

نحسب متوسط الوزن للمجموعتين بالكجم $\overline{x}_2$ و $\overline{x}_1$ على الترتيب

$$\overline{X}_1 = \frac{42+40+44+48+46}{5} = \frac{220}{5} = 44 \quad \text{كجم}$$

$$\overline{X}_2 = \frac{32+50+44+40+54}{5} = \frac{220}{5} = 44 \qquad \text{كجم}$$

مما سبق نجد أن $\overline{x}_1 = \overline{x}_2 = 44$ أي أن المجموعتين متساويتين في الوسط الحسابي مع أن المجموعة الأولى متجانسة وقيمها قريبة من المتوسط 44 بينما المجموعة الثانية غير متجانسة (متباعدة) وقيمها تبعد عن المتوسط 44 .

(١ , ٦ , ٢) المدى (Range)

يعرف المدى بأنه القيمة الناتجة من طرح أصغر قراءة من أكبر قراءة للبيانات محل الدراسة أي أن :

المدى (R) = أكبر قراءة - أصغر قراءة

مثال (٢ , ١٨)

احسب المدى لأوزان مجموعتي الطلاب في المثال السابق حيث:

المجموعة الأولى هي : 42, 40, 44, 48, 46

المجموعة الثانية هي : 32, 50, 44, 40, 54

المدى للمجموعة الأولى R_1 هو

$$R_1 = 48 - 40 = 8 \quad \text{كجم}$$

المدى للمجموعة الثانية R_2 هو

$$R_2 = 54 - 32 = 22 \quad \text{كجم}$$

أي أن R_1 أصغر من R_2 لتجانس البيانات في المجموعةالأولى أكثر من المجموعة الثانية .

(٢ , ٦ , ١ , ١) المدى في حالة الجداول التكرارية (البيانات المبوبة)

يوجد في هذه الحالة أكثر من تعريف ونكتفي بالتعريف التالي:

المدى = الحد الأعلى لأكبر فئة - الحد الأدنى لأصغر فئة .

مثال (٢ , ١٩)

احسب المدى لدرجات الطلاب من الجدول التالي :

فئات الدرجات	50 - 59	60 - 69	70 - 79	80 - 89	90 - 99
عدد الطلاب	2	2	9	11	6

الحل

المدى = الحد الأعلى لأكبر فئة - الحد الأدنى لأصغر فئة

درجة $99.5 - 49.5 = 50$

(٢ , ٦ , ١ , ٢) مميزات المدى

١ - مقياس سهل حسابه ويعطى فكرة سريعة عن تجانس البيانات .

٢ - يفيد في بعض التطبيقات مثل مراقبة ضبط جودة الإنتاج . وفي الأرصاد الجوية حيث يهمنا الحد الأعلى للحرارة والحد الأدنى لها .

(٢ , ٦ , ١ , ٣) عيوب المدى

١ - يتأثر كثيرا بالقيم المتطرفة نحو الصغر والكبر فهو مقياس تقريبي لا يعتمد عليه في التحليل الإحصائي .

٢ - يصعب حسابه في حالة الجداول التكرارية المفتوحة كما يصعب حسابه للبيانات الوصفية.

(٢ , ٦ , ٢) نصف المدى الربيعي (semi-interquartile Range)

للتغلب على القيم المتطرفة في الصغر أو الكبر عرف الإحصائيون المدى الربيعي في منطقة أكثر انتظاما في البيانات بعد ترتيبها تصاعديا أو تنازليا . حيث يحسب المدى الربيعي من الفرق بين الربيع الأعلى (Q_3) والربيع الأدنى (Q_1). يعرف نصف هذا الفرق بنصف المدى الربيعي (Q) ويحسب من العلاقة (٦ , ٢) التالية :

$$Q = \frac{Q_3 - Q_1}{2} \qquad (٢ , ٦)$$

مثال (٢٠ , ٢)

احسب نصف المدى الربيعي لأوزان الطلاب في مثال (١٨ , ٢) ثم قارن بين نصفي المدى الربيعي للمجموعة الأولى والثانية .

الحل

نصف المدى الربيعي Q للمجموعة الأولى حيث البيانات بعد ترتيبها هي :

40 , 42 , 44 , 46 , 48

الربيع الأول $Q_1 = 42$ كجم لأنها القيمة التي يسبقها $\frac{1}{4}$ البيانات .

الربع الثالث $Q_3 = 46$ كجم لأنها القيمة التي يسبقها $\frac{3}{4}$ البيانات .

ثم نحسب Q من العلاقة (٦ , ٢) كما يلي :

$$Q = \frac{46 - 42}{2} = 2 \text{ كجم}$$

ونصف المدى الربيعي Q^* للمجموعة الثانية حيث البيانات بعد ترتيبها هي :

32, 40, 44, 50, 54

فيكون الربيع الأول $Q_1^* = 40$ كجم والربيع الثالث $Q_3^* = 50$ ثم نحسب Q^* من العلاقة (٦ , ٢) كما يلي :

$$Q^* = \frac{Q_3^* - Q_1^*}{2} = \frac{50 - 40}{2} = 5 \text{ كجم}$$

ونلاحظ أن $Q = 2$ أقل من $Q^* = 5$ وهذا يعني أن المجموعة الثانية أكثر تشتتا من المجموعة الأولى .

(١ , ٢ , ٦ , ٢) نصف المدى الربيعي من الجداول التكرارية (البيانات المبوبة):

نحسب الربيع الأول Q_1 حيث يمثل المئين P_{25} ثم نحسب الربيع الثالث Q_3 حيث يمثل المئين P_{75} وذلك من العلاقة (٥ , ٢) ثم التعويض في العلاقة (٦ , ٢) لنحصل على قيمة نصف المدى الربيعي Q .

مثال (٢١ , ٢)

احسب قيمة نصف المدى الربيعي لدرجات الطلاب في مثال (٣ , ٢) السابق حسابيا وبيانيا :

الحل

نكون جدول التكرار المتجمع الصاعد لدرجات الطلاب ثم نحسب Q كما يلي :

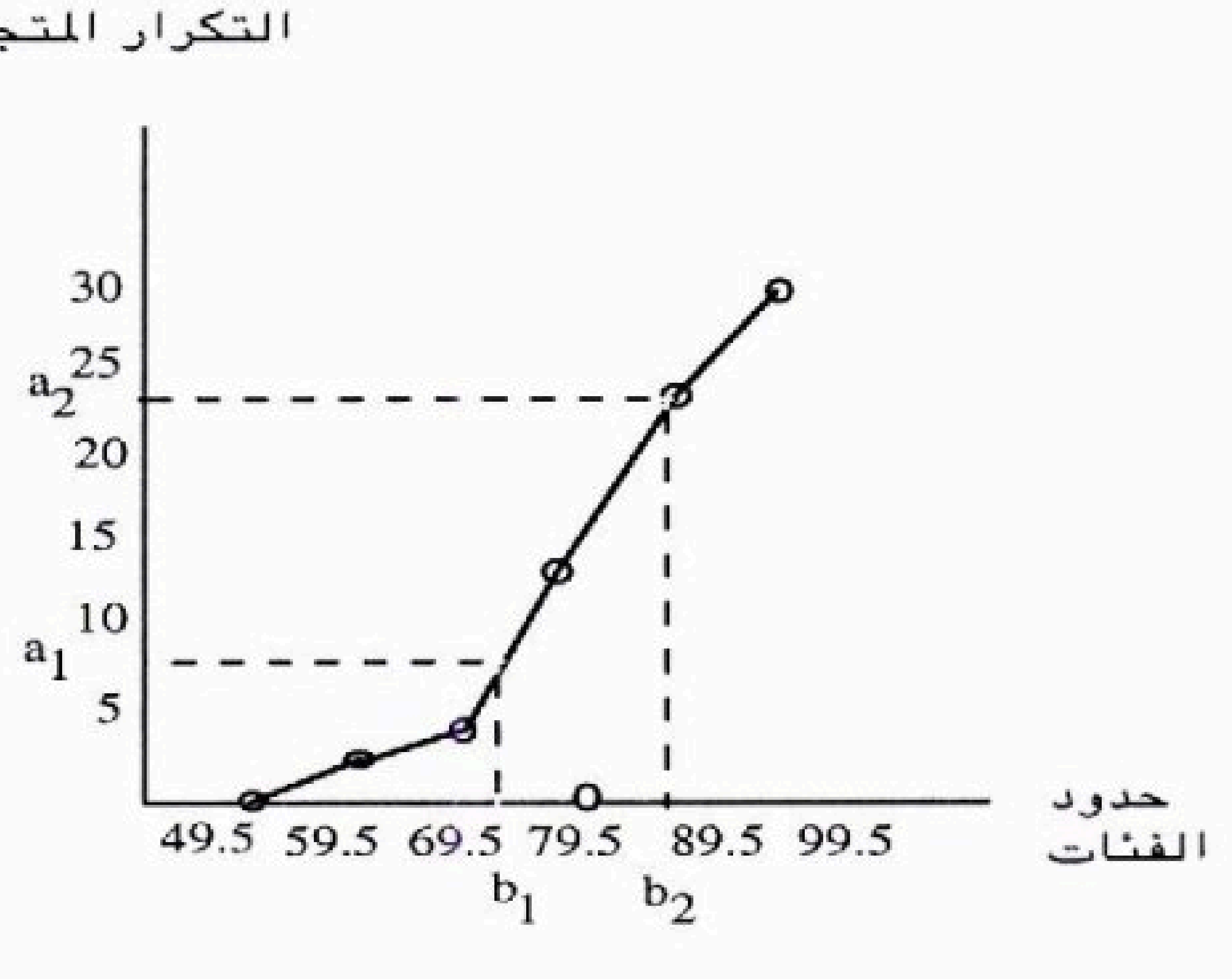
التكرار المتجمع الصاعد	حدود فئات الدرجات
0	أقل من 49.5
2	59.5 » »
4	69.5 » »
---------- Q_1	
13	79.5 » »
---------- Q_2	
24	89.5 » »
30	99.5 » »

$$Q_1 = P_{25} = 69.5 + \frac{7.5 - 4}{13 - 4} . 10 = 73.5$$ درجة

$$Q_3 = P_{75} = 79.5 + \frac{22.5 - 13}{24 - 13} . 10 = 88.1$$ درجة

$$Q = \frac{88.1 - 73.5}{2} = 7.3$$ درجة

ثم نحسب قيمة Q بيانيا من المضلع التكراري المتجمع الصاعد شكل (١٨ , ٢) كما يلي:



شكل (١٨ , ٢)

نحسب قيمة Q_1 بقراءة النقطة $b_1 = 73$ درجة تقريبا
نحسب قيمة Q_3 بقراءة النقطة $b_2 = 88$ درجة تقريبا

فيكون نصف المدى الربيعي (Q) بيانيا هو

$$Q = \frac{88 - 73}{2} = 7.5 \text{ درجة}$$

(٢ , ٢ , ٦ , ٢) مميزات وعيوب نصف المدى الربيعي:

هي مثل مميزات وعيوت الوسيط التي سبق ذكرها .

(٢ , ٦ , ٣) الانحراف المعياري (Standard deviation)

يعرف الإنحرف المعياري σ بأنه الجذر التربيعي للتباين σ^2 ويعرف التباين σ^2 بأنه متوسط مجموع مربعات انحرافات القراءات عن وسطها الحسابي . فإذا كان لدينا قراءات لمجموعة بيانات عددها n تمثل مجتمع ما هي :

$$x_1, x_2, ..., x_n$$

فإذا كان μ هو متوسط هذه القراءات فتكون انحرافات القراءات عن وسطها الحسابي μ هي :

$$(x_1 - \mu), (x_2 - \mu), ..., (x - \mu)$$

ويكون مربعات الانحرافات هو

$$(x_1 - \mu)^2, (x_2 - \mu)^2, ..., (x - \mu)^2$$

ويكون مجموع مربعات هذه الانحرافات هو $\sum_{i=1}^{n} (x_i - \mu)^2$ أي أن :

$$\sum_{i=1}^{n} (x_1 - \mu)^2 = (x_1 - \mu)^2 + (x_2 - \mu)^2 + \ldots + (x_n - \mu)^2$$

ويكون التباين σ^2 هو متوسط مجموع مربعات الانحراف هو

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} \tag{٢,٧}$$

حيث نستخدم كل قيم المجتمع

والانحراف المعياري σ هو

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}} \tag{٢,٨}$$

وعند دراسة مجموعة من البيانات كعينة من مجتمع ما وجد الإحصائيون أن التباين S^2 يعرف بأستخدام μ (القيمة الحقيقية لمتوسط المجتمع كما يلي :

$$S^2 = \Sigma(x_i - \mu)^2/n$$

أو إذا كانت μ غير معروفة فإننا نستخدم التقدير للتباين S^2 بمتوسط العينة $\overline{x}$ كما يلي :

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1} \tag{٢,٩}$$

والانحراف المعياري S هو

$$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \qquad (٢,١٠)$$

وبقليل من الاختصار الرياضي تصبح العلاقتين (٢,٩) و (٢,١٠) كما يلي :

$$S^2 = \frac{1}{n-1} (\Sigma x_i^2 - n \bar{x}^2) \qquad (٢,١١)$$

$$S = \sqrt{\frac{1}{n-1} (\Sigma x_i^2 - n \bar{x}^2)} \qquad (٢,١٢)$$

مثال (٢,٢٢)

احسب التباين S^2 والانحراف المعياري لأوزان الطلاب في مثال (٢,١٨) ثم قارن بين الأنحراف المعياري لكل من المجموعتين .

الحل

لحساب التبايـن S_1^2 ثم الانحراف المعياري S_1 للمجموعة الأولى نكـون الجدول التالي :

الوزن x	x^2
42	1764
40	1600
44	1936
48	2304
46	2116
Σ 220	9720

ثم نحسب S_1^2 و S_1 من العلاقتين (٢,١١) و (٢,١٢) كما يلي :

$$S_1^2 = \frac{1}{4}\,[9720 - 5(44)^2]$$

$$= \frac{1}{4}\,(9720 - 9680] = 10$$

$$S_1 = \sqrt{10} = 3.1$$

ولحساب التبايـن S_2^2 والانحراف المعياري S_2 للمجموعة الثانية نكـون الجـدول التالي ثم نستخدم العلاقتين (١١ , ٢) و (١٢ , ٢) كما يلي :

الوزن x	X^2
32	1024
50	2500
44	1936
40	1600
54	2916
$\Sigma x = 220$	$\Sigma x^2 = 10066$

$$S_2^2 = \frac{1}{4}\,[10066 - 5(44)^2]$$

$$= \frac{1}{4}\,[10066 - 9680] = 96.5$$

$$S_2 = \sqrt{96.5} = 8.1$$

نلاحظ الانحراف المعياري $S_1 = 3.1$ أقل مـن $S_2 = 8.1$ وهذا يعني أن المجموعة الأولى أكثر تجانسا من المجموعة الثانية أو المجموعة الثانية أكثر تشتتاً من المجموعة الأولى كما حدث ذلك أيضا بالنسبة لمقياس نصف المدى الربيعي في مثال (٢٠ , ٢) السابق .

ملاحظة

من خواص الانحراف المعياري نذكر خاصتين الأولى أن قيمته لا تتأثر بطرح

أو إضافة مقدار ثابت a . فمثلا إذا كان لدينا متغيرا $Y = X - a$ فإن $S_y = S_x$.
والثانية أن قيمته تتأثر بضرب مقدار ثابت c . فمثلا إذا كان لدينا متغيراً $Y = c\,X$ فإن $S_y = c\,S_x$.

مثال (٢,٢٣)

احسب الانحراف المعياري لأوزان مجموعة الطلاب للمرحلة الأولى في مثال (٢,٢٢) عندما يكون :

أولا : $Y = X - 40$ ثانيا : $Z = \frac{1}{2} X$.

الحل

سبق حساب $S_x = \sqrt{10} \approx 3.162$ للبيانات 42, 40, 44, 48, 46 والآن نكون جدولا لحساب S_y^2 ثم S_y كما يلي :

x الوزن	y = X - 40	Y^2
42	42 - 40 = 2	4
40	0	0
44	4	16
48	8	64
46	6	36
Σ	20	120

$$S_y^2 = \frac{1}{n-1}(\Sigma Y_i^2 - n\,\overline{Y}) = \frac{1}{4} = [120 - 5(16)]$$

$$= \frac{1}{4}[120 - 80] = 10$$

فيكون الانحراف المعياري كجم $S_y = \sqrt{10}$

وهو نفس الانحراف المعياري كجم $S_x = \sqrt{10}$

أي أن : $S_y = S_x$

ثانيا لحساب S_x نكون الجدول التالي :

x	$Z = \frac{1}{2}x$	Z^2
42	21	441
40	20	400
44	22	484
48	24	576
46	23	529
Σ	110	2430

$$S_z^2 = \frac{1}{n-1}(\Sigma Z^2 - n\overline{Z}^2)$$

$$= \frac{1}{4}[2430 - 5(22)^2]$$

$$= \frac{1}{4}[2430 - 2420]$$

$$= \frac{1}{4}[10]$$

$$S_z = \frac{1}{2}\sqrt{10}$$

أي أن: $S_z = \frac{1}{2} S_x$

أي أن الانحراف المعياري S_x لم يتأثر بطرح 40 من القراءات ولكن تأثر بضرب $\frac{1}{2}$ في القراءات مما يبين الخاصيتين السابقتين في الملاحظة السابقة .

(١ , ٣ , ٦ , ٢) التباين والانحراف المعياري من الجداول التكرارية (البيانات المبوبة)

يعرف التباين والانحراف المعياري بضرب التكرار f في بسط العلاقتين السابقتين (٩ , ٢) و (١٠ , ٢) حيث x يمثل مراكز الفئات وتصبح العلاقتان كما يلي:

$$S^2 = \frac{\Sigma f(x - \overline{x})^2}{n - 1} \qquad (٢ , ١٣)$$

$$S = \sqrt{\frac{\Sigma f(x - \overline{x})^2}{n - 1}} \qquad (٢ , ١٤)$$

وبقليل من الاختصار فإن العلاقتين (١٣ , ٢) و (١٤ , ٢) تصبحا كما يلي:

$$S^2 = \frac{1}{n-1}(\Sigma f_i x_i^2 - n\,\overline{x}^2) \qquad (٢ , ١٥)$$

$$S = \sqrt{\frac{1}{n-1}(\Sigma f_i x_i^2 - n\,\overline{x}^2)} \qquad (٢ , ١٦)$$

مثال (٢٤ , ٢)

احسب الانحراف المعياري لدرجات 30 طالبا في مثال (٣ , ٢) السابق.

الحل

لسهولة حساب S باستخدام العلاقتين (١٥ , ٢) و (١٦ , ٢) نكون الجدول التالي :

فئات الدرجات	التكرار f	مركز الفئات x	fx	fx^2
50 - 59	2	54.5	109	5940.5
60 - 69	2	64.5	129	8320.5
70 - 79	9	74.5	670.5	49952.25
80 - 89	11	84.5	929.5	78542.75
90 - 99	6	94.5	567.0	53581.50
	30		2405	196337.5

بالتعويض في العلاقة (٢,١٥) كما يلي :

$$\overline{X} = \frac{\Sigma f x}{n} = \frac{2405}{30} = 80.17 \quad \text{درجة}$$

$$S^2 = \frac{1}{29}[196337.5 - 30(80.17)^2]$$

$$= \frac{1}{29}[196337.5 - 192817] = \frac{1}{29}[3520.5] = 121.4$$

$$S = \sqrt{121.3} = 11 \quad \text{درجة تقريبا}$$

ملاحظة

يمكن تبسيط حسابات مثال (٢,٢٤) لدرجات الطلاب باستخدام الخاصية الأولى وهو طرح مقدار ثابت a من مراكز الفئات وليكن $a = 84.5$ يمثل مركز الفئة التي يناظرها أكبر تكرار حيث مراكز الفئات الجديد y هو $y = x - 84.5$ وبحساب S_y نجد أن S_y تساوي S_x كما يتضح من المثال (٢,٢٥) كما يلي (المقدار الثابت a يسمى وسط فرضي) :

مثال (٢,٢٥)

احسب الانحراف المعيـاري S في مثال (٢,٢٤) السابق باستخدام وسـط فرضي $a = 84.5$.

الحل

لسهولة الحل نكون جدولا للحل كما يلي :

فئة الدرجات	f	x	y = x - 84.5	f y	$f y^2$
50 - 59	2	54.5	54.5-84.5=-30	- 60	1800
60 - 69	2	64.5	-20	- 40	800
70 - 79	9	74.5	-10	- 90	900
80 - 89	11	84.5	0	0	0
90 - 99	6	94.5	10	60	600
	30		-50	- 130	4100

$$S_X^2 = S_Y^2 = \frac{1}{29}\left[4100 - 30\left(\frac{-130}{30}\right)^2\right]$$

$$= \frac{1}{29} = [4100 - 563] = \frac{1}{29}[3537] = 122$$

$$S = \sqrt{122} = 11 \quad \text{درجة تقريبا}$$

وهي نفس النتيجة للانحراف المعياري في مثال (٢,٢٤) السابق .

(٢,٦,٣,٢) مميزات وعيوب الانحراف المعياري

هي نفس مميزات وعيوب الوسط الحسابي التي سبق شرحها .

(٢,٦,٤) معامل الاختلاف (coefficient of variation)

سبق أن بينا أن المتوسط الحسابي $\overline{X}$ والانحراف المعياري S لهما وحدات قياس حسب طبيعة الظاهرة محل الدراسة - سم - كجم - سنة - درجة - . . . الخ وأحيانا يطلب مقارنة تجانس ظاهرتين طول مثلا بالسم ووزن مثلا بالكجم فإن المقارنة بالانحراف المعياري غير صحيحة لأن المقارنة تتم بوحدات مختلفة سم و كجم لذلك دعت الحاجة

إلى مقياس نسبي بدون وحدات هو معامل الاختلاف C.V ويعرف بالعلاقة (٢ , ١٧) كما يلي:

$$C.V = \frac{S}{\overline{X}} \qquad (٢ , ١٧)$$

مثال (٢ , ٢٦)

احسب معامل الاختلاف لدرجات الطلاب في مثال (٢ , ٢٥) السابق.

الحل

من مثال (٢ , ٢٥) نجد أن

$\overline{x} = 80.17$ درجة والانحراف المعياري $S = 11$ درجة فإن

$$C.V = \frac{S}{\overline{x}} = \frac{11}{84.17} = 0.14$$ تقريبا

(٢ , ٦ , ٥) القيمة المعيارية

أحيانا يتطلب المقارنة بين قراءتين أو أكثر ذات أوساط حسابية وانحرافات معيارية مختلفة استخدام ما يسمى بالقيم المعيارية حيث تتم المقارنة بدقة لهذه القيم. ويلاحظ أن القيم المعيارية أكثر بعداً من المركز (الصفر) تكون أفضل من القريبة من المركز ونعرف القيمة المعيارية Z كما يلي :

$$Z = \frac{X - \overline{X}}{S} \qquad (٢ , ١٨)$$

مثال (٢ , ٢٧)

أي درجتي أحد الطلاب أفضل في المقررين التاليين :

مقرر الفيزياء : درجة $X = 90 \ , \ \bar{x} = 80 \ , \ S_X = 5$

مقرر الكيمياء : درجة $Y = 85 \ , \ \bar{y} = 73 \ , \ S_Y = 4$

الحل

نحسب الدرجة المعيارية للفيزياء Z_1 والدرجة المعيارية للكمياء Z_2 كما يلي :

$$Z_1 = \frac{90 - 80}{5} = \frac{10}{5} = 2 \quad \text{درجة معيارية}$$

$$Z_2 = \frac{85 - 73}{4} = \frac{12}{4} = 3 \quad \text{درجة معيارية}$$

من الحسابات نجد أن $Z_2 = 3$ أكبر من $Z_1 = 2$ أي أن درجة الطالب في مادة الكيمياء أفضل من درجته في الفيزياء .

(٢,٦,٦) متباينة تشبشيف

أحيانا نرغب معرفة نسبة عدد البيانات داخل حدين يكون معلوم لدينا الوسط الحسابي $\bar{x}$ والانحراف المعياري s للبيانات لذلك نستخدم متباينة تشبشيف وهي: نسبة البيانات لمجموعة من البيانات عددها n داخل الحدين $(\bar{x} - ks \ , \ \bar{x} + ks)$ على الأقل هي $1-\frac{1}{k^2}$ حيث k مقدارا ثابتا موجبا أكبر من الواحد الصحيح $(k > 1)$.

مثال (٢,٢٨)

احسب حدود درجات الطلاب بحيث يكون داخلها 75% من عدد الطلاب على الأقل في مثال (٢,٢٦) السابق.

الحل

الحدود حسب متباينة تشبشيف هي :
$(\bar{x} - ks\ ,\ \bar{x} + ks)$ حيث درجة $\bar{x} = 80.17$ ، درجة $s = 11$ ولحساب k يتم كما يلي :

$$1 - \frac{1}{k^2} = 0.75 \quad \Rightarrow \quad \frac{1}{k^2} = 1 - 0.75 = 0.25$$

$$\Rightarrow \quad k^2 = 4 \qquad \Rightarrow \quad k = 2$$

بالتعويض عن $k,\ \bar{x},\ s$ داخل حدود متباينة تشبشيف نجد أن :

$$(80.17 - 2(11)\ ,\ \ 80.17 + 2(11))$$

أي أن

$$(58.17\ ,\ 102.17)$$

(٢,٧) مقاييس الالتواء والتفلطح

على مسار وصف البيانات تكلمنا عن المنحنيات التكرارية المتماثلة منـهـا وغير المتماثلة ولقياس بعدها عن التماثل أو قربها نعرف معامل الالتواء ولقيـاس ارتفاع قمتها أي تدببها أو انخفاض قمتها أي تفلطحها نعرف معامل التفـلـطـح وذلك كما يلي:

(٢,٧,١) معامل الالتواء

يعرف معامل الالتواء بقسمة العزم الثالث m_3 للبيانات على مكعب الانحراف المعياري S ونرمز له بالرمز γ ويعطى كما يلي :

$$\gamma = \frac{m_3}{S^3} \qquad (٢,١٩)$$

حيث $m_3 = \frac{\Sigma(x - \overline{x})^3}{n}$ للبيانات المباشرة التي حجمها n .

و $m_3 = \frac{\Sigma f(x - \overline{x})^3}{n}$ للبيانات من الجداول التكرارية.

(٢,٧,٢) معامل التفلطح

يعرف معامل التفلطح بقسمة العزم الرابع m_4 على الانحراف المعياري مرفوعا للقوى الرابعة (m^4) ويرمز له بالرمز K_r ويعطى بالعلاقة التالية :

$$K_r = \frac{m_4}{s^4} \tag{٢,٢٠}$$

حيث $m_4 = \frac{\Sigma(x - \overline{x})^4}{n}$ للبيانات المباشرة .

و $m_4 = \frac{\Sigma f(x - \overline{x})^4}{n}$ للبيانات من الجداول التكرارية .

(٢,٨) تمارين

١ - أعطي بعض الأمثلة عن ظواهر بياناتها عبارة عن:
وصفية - وصفية ذات تراتيب - بيانات كمية ذات مدى صغير نسبيا - بيانات كمية ذات مدى كبير نسبيا.

٢ - فيما يلي بيانات عن لون العينين 25 شخصا تم اختيارهم عشوائيا من أحد المجتمعات .

أسود	أزرق	بني	أخضر	أسود
أسود	أسود	أزرق	أسود	بني
أخضر	بنى	أسود	أزرق	أخضر
بني	أزرق	بني	بني	أسود
أسود	أخضر	بني	أسود	أسود

ا) كون جدولا تكراريا لهذه البيانات.

ب) مثل هذه البيانات بواسطة الأعمدة.

٣ - البيانات التالية تمثل عدد الساعات التي عملهـا 40 عاملا في مصنع فـي أسبوع معين

49	36	41	43	38	32	24	45
23	41	43	48	46	34	24	46
23	48	47	33	36	22	39	44
46	37	39	41	26	28	33	43
25	35	36	43	48	47	38	43

أ) كون جدولا تكراريا ذو خمس فئات .

ب) ارسم المدرج التكراري والمضلع التكراري والمنحنى التكراري لهذا الجدول.

ج) أوجد التكرار النسبي لهذا الجدول وارسم مضلعه التكراري .

د) أوجد التكرار المتجمع الصاعد وارسم مضلعه التكراري .

هـ) احسب الوسط الحسابي - المنوال - الوسيط لهذا الجدول التكراري .

و) احسب معامل الاختلاف لهذا الجدول التكراري .

ز) أوجد حدود متباينة تشبيشيف عندما $K = 2$ لهذه البيانات

ح) احسب معامل الالتواء ومعامل التفلطح لهذه البيانات .

٤ - سجلت أطوال 40 ورقة من أوراق نبات الغار لأقرب سم:

15	14	14	15	14	15	14	15
15	14	14	14	17	14	16	15
16	13	13	16	16	14	15	15
13	18	12	17	16	15	15	14
15	14	16	13	17	15	15	14

ا) كون توزيعا تكراريا مناسبا .

ب) ارسم المدرج التكراري والمضلع التكراري النسبي .

ج) أوجد الجدول التكراري المتجمع الصاعد والهابط وارسم كل من المضلع المتجمع الصاعد والمضلع المتجمع الهابط.

د) احسب المنوال والوسيط والمئين التسعين (P_{90}) لهذه البيانات .

هـ) احسب المدى - ونصف المدى الربيعي من الجدول التكراري.

٥ - فيما يلي توزيع الأجر اليومي لعدد من العمال بالريال في أحد المصانع:

فئات الأجر	30-39	40-49	50-59	60-69	80-79	80-89
عدد العمال	9	10 .	15	8	4	2

ا) احسب الوسط الحسابي لأجور العمال .

ب) أوجد الوسيط والمنوال حسابيا وبيانيا .

ج) إذا كان الأجر اليومي لكل عامل يزيد بمقدار خمسة ريالات كل ستة شهور فما قيمة المقاييس السابقة بعد سنة ؟

٦ - إذا كانت أسعار ثلاثة أنواع من الفاكهة 26, 36, 41 ريالا على التوالي للصندوق . إذا باع تاجر 50 صندوقا من النوع الأول ، 30 صندوقا من النوع الثاني و 20 صندوقا من الثالث .

أوجد متوسط سعر البيع للصندوق الواحد .

٧ - إذا كانت الحمولة القصوى لمصعد إحدى المؤسسات 3000 رطلا فهل تعد الحمولات التالية أكبر من طاقة المصعد .

أ) إذا صعد 14 عاملا وزن كل منهم 160 رطلا .

ب) إذا صعد 11 عاملا وزن كل منهم 135 رطلا وعشرة آخرين وزن كل منهم 165 رطلا .

٨ - عند دراسة ظاهرة الطول والوزن لمجموعة طلاب بإحدى الكليات كانت لدينا البيانات التالية :

ا) ظاهرة الطول $\overline{x} = 165 \text{ cm} \quad , \quad s = 9 \text{ cm}$

ب) ظاهرة الوزن $\sum_{i=1}^{20} x_i = 1200 \text{ kg} \,, \quad \sum_{i=1}^{20} x_i^2 = 7287 \text{ kg}$

أي من الظاهرتين أكثر تجانسا .

٩ - في دراسة قام بها مركز للأغذية وجد أن متوسط كمية الفيتامين B في شرائح الخبز هي 0.25 ملجرام بانحراف معياري 0.005 مليجرام .

ا) أوجد حدود الفترة التي تقع بينها كمية فيتامين B في نسبة $\frac{35}{36}$ من الشرائح على الأقل .

ب) فما هي أقل نسبة من الشرائح التي تحتوى على مقدار من هذا الفيتامين واقع بين (0.232, 0.248) .

١٠ - حول هذه القيم التالية 5, 4, 6, 1, 2, 8 إلى قيم معيارية ثم أوجد متوسطها وتباينها .

١١ - احسب المتوسط $\overline{x}$ والانحراف المعياري s والعزم الثالث m_3 والعزم الرابع m_4 ثم احسب معامل الالتواء γ ومعامل التفلطح K_r للبيانات التالية .

2, 5, 6, 8, 9, 11

١٢ - في دراسة عن أحد النباتات التي لها نفس العمر كانت أطوالها بالسم كما في الجدول التالي :

فئات الطول	25-29	30-34	35-39	40-44	45-49
عدد النباتات	14	52	50	36	8

ا) احسب معامل الاختلاف لهذه البيانات .

ب) احسب معامل الالتواء ومعامل التفلطح لهذه البيانات .

١٣ - فيما يلي تصنيف لعدد أيام الغياب خلال فصل دراسي لشعبة تتضمن 46 طالبا .

فئات الطول	0	1	2	3	4
عدد النباتات	20	10	8	5	3

احسب المنوال والوسيط والمتوسط ونصف المدى الربيعي لعدد أيام الغياب في هذه الشعبة .

١٤ - ادعت شركة طيران أن سفرياتها بين المدن الداخلية تصل متأخرة عن موعدها بمتوسط 4.5 دقيقة وانحراف معياري 1.6 دقيقة . فما هي أقل نسبة من سفرياتها تصل متأخرة ما بين (1.8 و 1.4 دقيقة).

١٥ - باستخدام القيم التالية 6, 5, 7,2,3,9 أثبت أن متوسط مجموعة الدرجات المعيارية هو صفر وانحرافها المعياري هو واحد صحيح.

الفصل الثالث

مبادىء الاحتمالات
Principles of Probabilities

● مقدمة ● المجموعـات ● طرق العد ● التجربة العشوائية وفراغ العيـنـة والحـدث ● االتعريـف الكلاسيكي للاحـتـمـال ● التعريف التجريـبـي للاحتمال ● التعريف الحديث للاحتمال (مسلمات الاحتمال) ● الاحتمال الشرطي والاستقـــلال ● الاحتمال الكلي ● نظرية بيز ● تمارين

(١ , ٣) مقدمة

ينقسـم حـدوث الظـواهر الطبيعيـة أو غـيرها في حيـاتنا اليوميـــة إلـى نوعــين :

النوع الأول : حوادث مؤكـدة (certain events) مثل شروق الشمس صباحـا مـن المشرق يوميا . سقوط حجر عند القائه من أعلى منزل إلى سطح الأرض . . . الخ وهـذا النـوع يسمى النموذج المؤكد (deteminstic model) .

النوع الثاني : حوادث غير مؤكدة وتعتمـد على الصـدفــة (chance events) مثل سقوط المطر في يوم تكون فيه السماء ملبدة بالغيـوم ، وأن يـكـون المـصـبـاح محروقا من إنتاج أحد المصانع ، أو فوز فريق كرة قدم على فريق آخر ، أو وصول طائرة الخطوط السعودية في ميعادها . . . الخ . وأحيانا نجد أننا نعبر عن هـذه الاحتمالات بتقدير عددي . فمثلا نقول احتمال سقوط المطـر 25% . واحتمال

وصول طائرة الخطوط السعودية في ميعادهاإلى مطار الملك خالد الدولي بالرياض 98%. ومثل هذه الحوادث التي تعتمد على معلومات غير مؤكدة أو الصدفة تسمى بالنموذج الاحتمالي (probablistic model) . وعلم الاحتمالات يعتمد على ذلك النموذج الاحتمالي وهو مثل العلوم الأخرى يبدأ ببعض التعاريف والمسلمات نستعرض بعضها كما يلي :

(٣ , ٢) المجموعات (Sets)

تعرف المجموعة بأنها تجميع أي عدد من العناصر تجمعها صفة مشتركة ويرمز للمجموعات بالحروف الكبيرة ....A,B,C وعناصر المجموعة بالحروف الصغيرة ...a,b,c ويستخدم الرمز $\in$ ليمثل الانتماء ، فمثلا $b \in B$ يعني أن العنصر b ينتمي للمجموعة B والرمز $\notin$ يمثل عدم الانتماء ، فمثلا $b \notin C$ يعني أن العنصر b لا ينتمي للمجموعة C .

(٣ , ٢ , ١) تمثيل المجموعات

وتمثل المجموعة بطريقتين الأولى: طريقة القائمة. والثانية : طريقة الصفة المميزة ونوضح ذلك كما يلي:

أولا: طريقة القائمة (roster method)

هي أن توضع جميع العناصر بين قوسين والمجموعات التالية أمثلة على ذلك:

$A = \{h,t\}$ ، $B = \{1,2,3,4,5,6,\}$ ، $S = \{ab, ac, bc\}$ . . الخ .

ثانيا: طريقة الصفة المميزة (rule method)

هي أن توضع الصفة المميزة بين قوسين مثل $A = \{ x \mid$ عدد زوجي $x \}$.

(٣ , ٢ , ٢) أنواع المجموعات

المجموعة الخالية (empty set)

هي المجموعة التي لا تشتمل على أية عناصر ويرمز لها بالرمز ϕ أو $\{\ \}$.

المجموعة الشاملة (universal set)

هي المجموعة التي تشتمل على جميع عناصر جميع المجموعات الجزئية التي تعتبر مجموعات جزئية منها.

المجموعة الجزئية (subset)

نقول إن المجموعة B مجموعة جزئية من D ويرمز لها $B \subset D$ حيث الرمز $\subset$ يعني مجموعة جزئية من والرمز $\not\subset$ يعني ليست مجموعة جزئية من . فمثلا المجموعة A ليست جزئية من D تكتب $A \not\subset D$.

المجموعتين المتساويتين (equality)

يقال إن المجموعة B تساوي المجموعة D إذا كانت B مجموعة جزئية من D وأن D مجموعة جزئية من B أي أن :

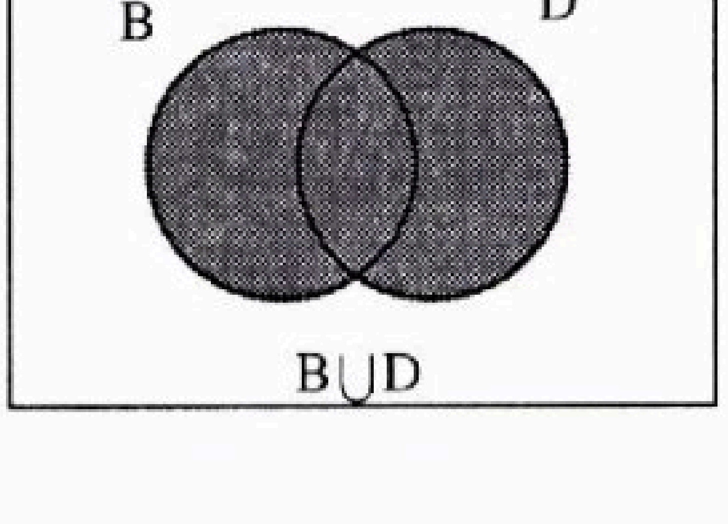
$$B \subset D, D \subset B \Rightarrow B = D$$

(٣ , ٢ , ٣) العمليات الجبرية على المجموعات (Set operations)

من العمليات الجبرية على المجموعات الاتحاد والتقاطع والمجموعة المكملة وتساوي مجموعتين ونوضح ذلك كما يلي :

(١ , ٣ , ٢ , ٣) الاتحاد (union)

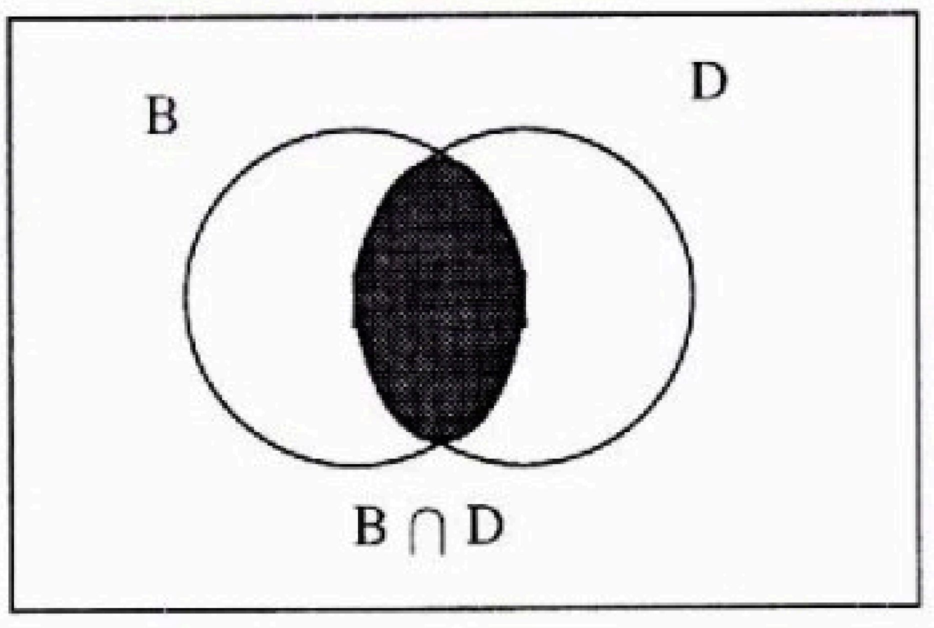
اتحاد مجموعتين B و D ويرمز بالرمز $B \cup D$ وهو عبارة عن مجموعة العناصر الموجودة في B أو D أو كليهما ويكتب رياضيا كما يلي: $B \cup D = \{ x \mid x \in B$ أو $x \in D \}$ ، ويوضح في شكل فن بالجزء المظلل شكل (١ , ٣) كما يلي:



شكل (١ , ٣)

(٢ , ٣ , ٢ , ٣) التقاطع (intersection)

يعرف التقاطع بين المجموعتين D و B ويرمز له بالرمز $B \cap D$ وهو عبارة عن مجموعة العناصر المشتركة بين D و B وبعبارة أخرى $B \cap D = \{x \mid x \in D,\ x \in B\}$. يوضح التقاطع بالجزء المظلل في شكل فن (٢ , ٣) كما يلي :



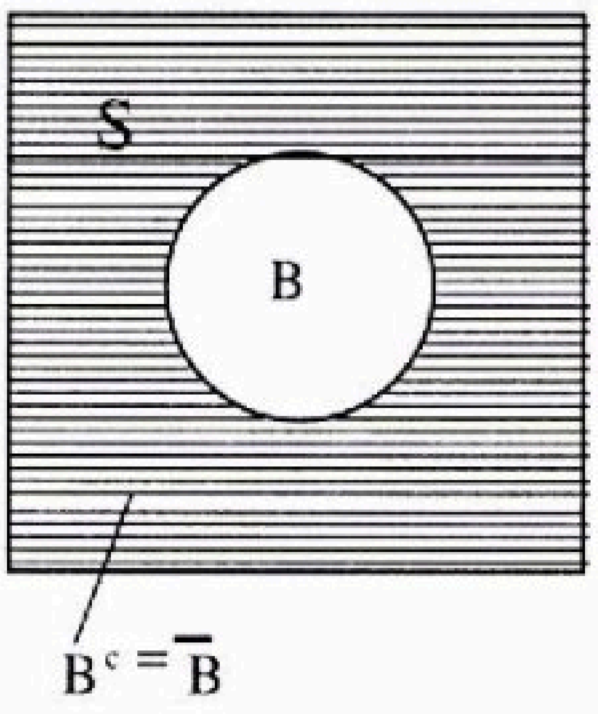
شكل (٢ , ٣)

(٣ , ٣ , ٢ , ٣) الفرق بين مجموعتين (diference)

يعرف الفرق $A - B$ بأنه المجموعة التي تضم العناصر الموجودة في A وليست موجودة في B ويكتب أحيانا $A \cap \overline{B}$.

(٤ , ٣ , ٢ , ٣) المجموعة المتممة (complement set)

المجموعة المتممة للمجموعة B هي مجموعة العناصر للمجموعة الشاملة S والتي لا تنتمي للمجموعة B ويرمز لها بالرمز $\overline{B}$ أو B^c وبعبارة أخرى $B^c = \overline{B} = \{x \mid x \in S,\ x \notin B\}$ وتوضح في شكل فن بالجزء المظلل شكل (٣ , ٣) كما يلي :



شكل (٣,٣)

(٦, ٣, ٢, ٣) نذكر باختصار بعض العمليات الجبرية على المجموعات

(i) خاصية التبديل (commutative law)

التبديل على الاتحاد $B \cup D = D \cup B$. التبديل على التقاطع $B \cap D = D \cap B$.

(ii) خاصية التجميع (associative law)

التجميع على الاتحاد $B \cup (D \cup C) = (B \cup D) \cup C$

التجميع على التقاطع $B \cap (D \cap C) = (B \cap D) \cap C$

(iii) خاصية التوزيع (distributive law)

توزيع التقاطع على الاتحاد $B \cap (D \cup C) = (B \cap D) \cup (B \cap C)$

توزيع الاتحاد على التقاطع $B \cup (D \cap C) = (B \cup D) \cap (B \cup C)$

(v) قوانين دي مورجان (de Morgan's laws)

$$\overline{B \cup D} = \overline{B} \cap \overline{D}$$

$$\overline{B \cap D} = \overline{B} \cup \overline{D}$$

ملاحظة

القوانين السابقة يمكن أن تعمم لأكثر من مجموعتين.

(٣,٣) طرق العَدّ Counting

هي طرق رياضية تساعدنا في إيجاد عدد عناصر المجموعة دون الحاجة إلى كتابة عناصر المجموعة سواء بجدولة العناصر أو الصفة المميزة وهذه الطرق مفيدة في حساب الاحتمالات كما سوف يتضح فيما بعد . نستعرض منها :

ا - قاعدة الضرب .

ب - قاعدة الجمع .

ج - التباديل .

د - التوافيق .

هـ - التباديل داخل أشياء متساوية .

(١,٣,٣) قاعدة الضرب (Rule of multiplication)

إذا كان لدينا عملية تتم على مرحلتين المرحلة الأولى تتم بعدد طـرق n_1 والثانية بعدد طرق n_2 ، فإن عدد طرق العملية في المرحلتين معا هو $n_1\ n_2$.

مثال (١,٣)

بكم طريقة يمكن أن يختار مسافر إلى الدمام من الرياض ثم يسـافـر مـن الدمام إلى الجبيل علما أنه يوجد من الرياض إلى الدمام ثلاث طرق تناسبه ومن الدمام إلى الجبيل طريقتان تناسبة أيضا .

الحـــل

من الشجرة البيانية فيما يلي يتضح أن عدد طرق السفر من الريـاض إلــى الجبيل تساوي 6 طرق ، وحسب قاعدة الضرب فإن :

عدد طرق السفر من الرياض إلى الدمام = 3 طرق

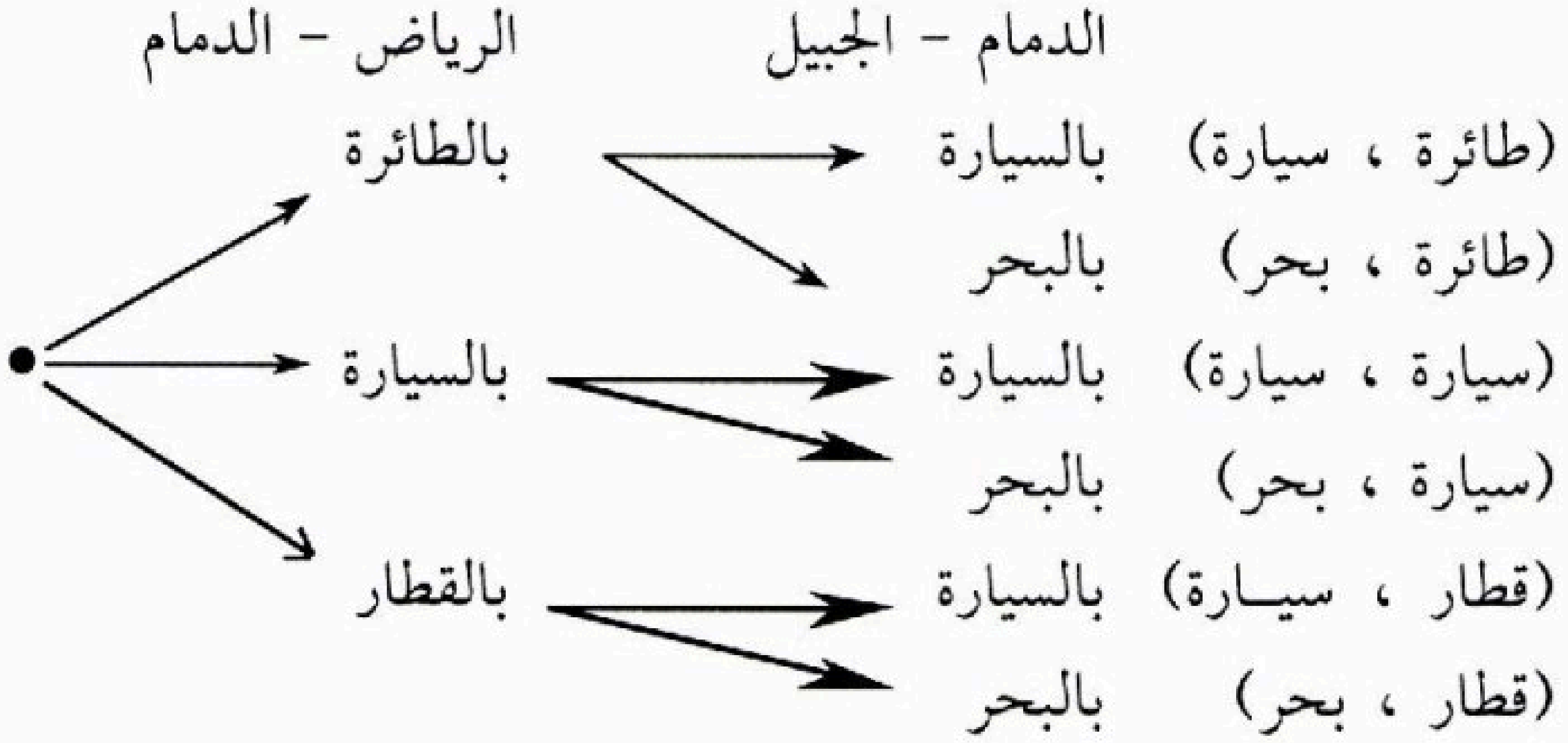
عدد طرق السفر من الدمام إلى الجبيل = 2 طريقة .

وحسب قاعدة الضرب عدد طرق السفر من الريـاض إلى الجبيـل = 2×3 = 6 طرق.

وهي نفس النتيجة من الشجرة البيانية ،

ويمكن تعميم قاعدة الضرب إلى أكثر من مرحلتين .

نستعين بالشجرة البيانية التالية لتوضيح الاختيارات الممكنة كما يلي:



(٢ , ٣ , ٣) قاعدة الجمع (Addition rule)

إذا كان لدينا عمليتين متنافيتين D و B بحيث تتم العملية B بعدد طرق n_1 والعملية D بعدد طرق n_2 فإن عدد طرق إتمام B أو D يساوي $n_1 + n_2$.

مثال (٣ , ٢)

طالب سعودي حاصل على شهادة الثانوية يريد أن يلتحق بجامعة الملك سعود بالرياض أو جامعة الملك عبد العزيز بجدة وكان مقبولا في جامعة الملك سعود في أربع كليات وجامعة الملك عبد العزيز في خمس كليات . ما هو عدد طرق قبوله في إحدى الكليات بالجامعتين السابقتين.

الحل

يتم عدد الاختيارات المختلفة كما يلي:

عدد طرق قبوله بجامعة الملك سعود =٤ كليات

عدد طرق قبوله بجامعة الملك عبدالعزيز =٥ كليات

حسب قاعدة الجمع فإن:

عدد طرق قبوله في إحدى الكليات بالجامعهين = ٤+٥=٩ طرق.

وبمكن تعميم قاعدة الجمع لأكثر من عمليتين متنافيتين (أو مانعتين).

(٣,٣,٣) التباديل (Permutation)

هي ترتيبة لعدة أشياء مختلفة بأخذها كلها أو بعضها في كل مرة مع مراعاة الترتيب . وعدد تباديل n من الأشياء مأخوذة في كل مرة r من الأشياء المختلفة (حيث $r \leq n$) في كل مرة يساوي عدد الترتيبات التي يمكن تكوينها من هذه الأشياء بحيث يحوي كل ترتيبة منها r من الأشياء المختلفة .

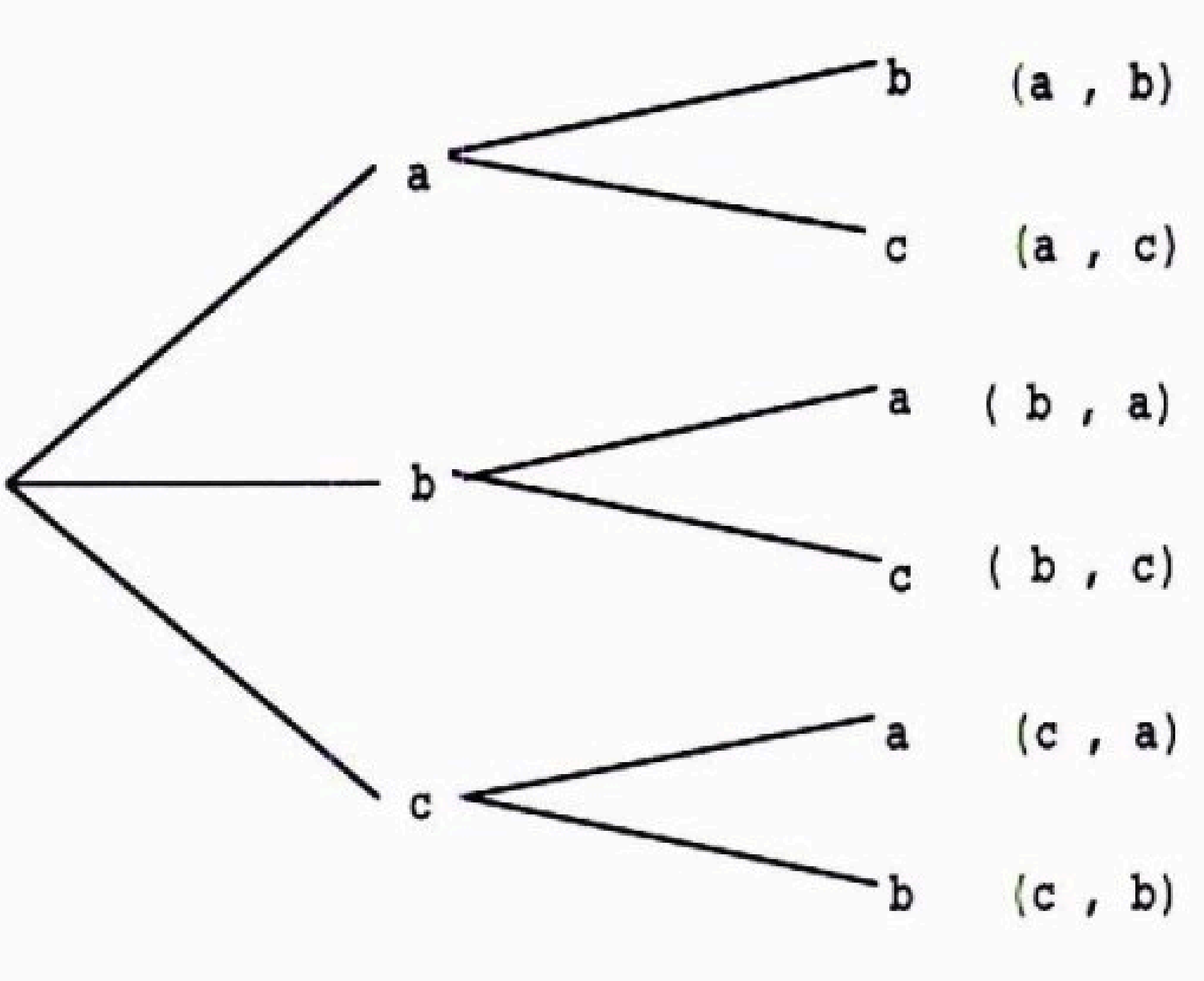
مثال (٣,٣)

أوجد عدد تباديل حرفين من الثلاثة حروف a, b, c بحيث لا يتكرر أي حرف في أي تبديلة .

الحل

نرسم الشجرة البيانية للاختبارات المختلفة كما يلي :

اختيار الحرف الأول   اختيار الحرف الثاني



فإن عدد تباديل حرفين من ثلاثة = 6 طرق

وحسب قاعدة الضرب فإن :

عدد طرق اختيار الحرف الأول = 3 طرق

عدد طرق اختيار الحرف الثاني = 2 طريقة

عدد طرق اختيار الحرفين معا = $3(2) = 6$ طرق .

وفي مثال (٣,٣) يمكن أن نرمز لعدد تباديل حرفين مختلفين من ثـلاثـة حروف مختلفة بالرمز $^{3}P_{2}$. حيث إن $^{3}P_{2} = 3\times 2$ وبوجه عام يرمز لعدد تباديل r من الأشياء المختلفة من n من الأشياء المختلفة $(r \le n)$ بالرمز $^{n}P_{r}$ حيث إن :

$$^{n}P_{r} = n(n-1)(n-2)\cdots(n-r+1) \qquad (٣,١)$$

وعندما $r = n$ فإن العلاقة (١,٣) تصبح $^{n}P_{n}$ كما يلي :

$$^{n}P_{n} = n(n-1)(n-2)\cdots 3\times 2\times 1 \qquad (٣,٢)$$

ويسمى المقدار $^{n}P_{n}$ وهو اختيار n شيء مختلف مـن n شيء مختلف بمضروب n ويرمز له بالرمر $n!$. وتصبح العلاقة (٢,٣) كما يلي :

$$n! = n(n-1)(n-2)\cdots 3\times 2\times 1 \qquad (٣,٣)$$

ويمكن إعادة كتابة (١,٣) بصورة مختصرة كما يلي :

$$^{n}P_{n} = \frac{n!}{(n-r)!} \qquad (٣,٤)$$

حيث يمكن كتابة العلاقة (٤,٣) كما يلي :

$$^{n}P_{n} = \frac{n(n-1)(n-2)\cdots(n-r+1)(n-r)(n-r-1)\cdots 3\times 2\times 1}{(n-r)(n-r-1)\cdots 3\times 2\times 1}$$

أي أن

$$ {}^{n}P_{r} = n(n-1)\cdots(n-r+1) $$

وهي نفس العلاقة (١ , ٣) السابقة .

مثال (٤ , ٣)

أوجد قيم $5!,\ {}^{6}P_{2},\ 0!$

الحل

$$ 5! = 5\times4\times3\times2\times1 = 120 $$

$$ {}^{6}P_{2} = 6\times5 = 30 $$

من العلاقة (٤ , ٣) نضع $r = n$ لنحصل على :

$$ {}^{n}P_{n} = \frac{n!}{(n-n)!} $$

أي أن

$$ {}^{n}P_{n} = \frac{n!}{0!} $$

أي أن

$$ 0! = \frac{n!}{{}^{n}P_{n}} = 1 $$

فإن مضروب صفر هو الواحد الصحيح $(0! = 1)$.

ملاحظات

١) السحب بدون إرجاع لعدد r شيء مختلف من n شيء مختلف يعطي بالتباديل ${}^{n}P_{r}$ كما هو مبين في العلاقة (٤ , ٣) السابقة.

٢) السحب بارجاع يكون عدد التباديل n^r ويوضح كما يلي :

عدد طرق اختيار العنصر الأول من n شيء مختلف =n طريقة

عدد طرق اختيار العنصر الثاني من n شيء مختلف = n طريقة .

.

.

عدد طرق اختيار العنصر r من n شيء مختلف = n طريقة .

عدد اختيار r شيء من n شيء مختلف بارجاع $(r \le n)$ حسب قاعدة الضرب =

$$n^r = \underbrace{n \times \ldots n \times n}_{r \text{ مرة}}$$ طريقة .

مثال (٣ , ٥)

وعاء به 20 كرة مرقمة من 1 إلى 20 ما هو عدد الطرق لسحب كرتين :

أولا : بدون إرجاع ثانيا : بإرجاع

الحل :

أولا : عدد طرق سحب كرتين بدون ارجاع هو

$^{20}P_2 = 20 \times 19 = 380$ طريقة

ثانيا : عدد طرق سحب كرتين بارجاع هو

$20^2 = 20 \times 20 = 400$ طريقة

(٣ , ٣ , ٤) التوافيق (Combinations)

هي اختيار r شيء مختلف من n شيء مختلف $(r \le n)$ مع عدم الأخذ في الاعتبار الترتيب ويرمز لها ^nC_r أو $\binom{n}{r}$ وتوجد علاقة بين التوافيق والتباديل هي :

$$\binom{n}{r} = {}^nC_r = \frac{^nP_r}{r!} = \frac{n!}{r!(n-r)!} \qquad (٣ , ٥)$$

مثال (٣,٦)

أوجد قيم التوافيق التالية :

$$\binom{5}{5}, \binom{6}{4}, \binom{4}{0}$$

الحل

$$\binom{5}{5} = \frac{5\,!}{5!(5-5)!} = \frac{5\,!}{5!\,0!} = 1$$

$$\binom{6}{4} = \frac{6!}{4!(6-4)!} = \frac{6!}{4!\,2!} = \frac{6\times5\times4!}{4!\,2\times1} = 15$$

$$\binom{4}{0} = \frac{4!}{0!(4-0)!} = \frac{4!}{0!\,4!} = 1$$

ملاحظة :

توجد علاقة بين اختيار r شيء مختلف من n شيء مختلف واختيار (n-r) شيء مختلف من n شيء مختلف هي :

$$\binom{n}{r} = \binom{n}{n-r}$$

(٣,٣,٥) التباديل داخل أشياء متساوية

نفرض أن لدينا n من الاشياء تحتوي على r من المجموعات كل مجموعة يوجد فيها الأشياء نفسها فإذا كانت المجموعات يوجد بها الأشياء نفسها هي $n_1, n_2, \ldots, n_r$. فإن عدد التبديلات حسب قاعدة الضرب هي :

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1\cdots n_{r-2}}{n_{r-1}} = \frac{n!}{n_1!n_2!\cdots n_r!} \qquad (٣,٦)$$

مثال (٣,٧)

بكم طريقة يمكن ترتيب حروف كلمة Statistics .

الحل

عدد تكرار حرف S هي $n_1 = 3$ مرات

عدد تكرار حرف t هي $n_2 = 3$ مرات

عدد تكرار حرف a هي $n_3 = 1$ مرة

عدد تكرار حرف i هي $n_4 = 2$ مرة

عدد تكرار حرف c هي $n_5 = 1$ مرة

فإن عدد تراتيب حروف كلمة Statistics $= \dfrac{10!}{3!\ 3!\ 1!\ 2!\ 1!}$

$= 50400$ مرة

(٣,٤) التجربة العشوائية وفراغ العينة والحدث

(Random experiment, Sample space and Event)

(٣,٤,١) التجربة العشوائية (Ranodm experiment)

هي إجراء نعلم مسبقا جميع النتائج الممكنة له وإن كنا لا نستطيع أن نتنبأ أي هذه النتائج سيتحقق فعلا . فمثلا عند رمي قطعة عملة نعلم ظهور الوجه الأعلى إما صورة أو كتابة ولكن لا ندري من سيظهر أولا . كذلك عند رمي حجر النرد فإن عملية الرمي هي تجربة عشوائية نعلم ظهور الوجه الأعلى أحد الأرقام 1,2,3,4,5,6 ولكن لا نعلم أيهم يحدث أولاً .

(٣,٤,٢) فراغ العينة (Sample space)

لكل تجربة عشوائية عدد من النتائج الممكنة . مجموعة هذه النتائج الممكنة

يسمى فراغ العينة ويرمز لها بالرمز S . ففي حالة قطعة العملة إذا رمزنا لظهور الصورة بالرمز H ولظهور الكتابة بالرمز T فيكون فراغ العينة لرمي قطعة عملـة مـرة واحدة هـو $S = \{H,T\}$ ويكون فراغ العينة لرمي حجـر نرد مـرة واحدة هـو $S = \{1,2,3,4,5,6\}$ وهكذا .

(٣ , ٤ , ٣) الحدث (Event)

أحيانا يكون اهتمامنا منصبـاً على نتيجة معينة من نتائج التجربة العشوائية وتسمى حدث أو حادثة وهو عبارة عن مجموعة جزئية من فراغ العينة . يرمز عادة للحدث بالرموز الكبيرة ...,A,B,C. وعناصرها بالرموز الصغيرة أحيانـا ...,a,b,c فإن حدث ظهور الصـورة هو $A = \{H\}$ عند رمي قطعة عملة مرة واحدة . فإن حدث ظهور رقم زوجي عند رمي حجر النرد مرة واحدة هو $B = \{2,4,6\}$. أي أن الحدث يمكن أن يكون بسيطا يتكون من عنصرا واحدا مثل ظهور الصورة عند رمي قطعة العملة مرة واحدة، أو يكون أكثر من عنصر وذلك في الحدث B لظهور عدد زوجي عند رمي حجر النرد مرة واحدة .

(٣ , ٤ , ٤) معنى بعض الحوادث

(i) اتحاد الحادثتين D و B أي $B \cup D$.

تتكون الحادثة $B \cup D$ من نقاط (عناصر) فراغ العينـة S والموجودة إما في الحادثة B أو الحادثة D أو كليهما وبمعنى آخر حدوث الحادثة $B \cup D$ يعني حدوث أحدهما على الأقل (أي حدوث أحدهما أو كليهما) .

(ii) تقاطع الحادثتين D و B أي $B \cap D$ أو تكتب BD .

تتكون الحادثة $B \cap D$ من نقاط العينة S والموجودة في كل من B و D معا. ويلاحظ أن الحادثتين المتنافيتين (المنفصلتين) تقاطعهما هو $B \cap D = \phi$.

(iii) الحادثة المكملة $\overline{B}$ (B^c)

الحادثة المكملة $\overline{B}$ تتكون من نقاط العينة S وغير الموجودة في الحدث B . يعني وقوع الحدث $\overline{B}$ يعني عدم وقوع الحدث B ويكتب $\overline{B}$. وأن الحدثين $\overline{B}$ و B متنافيين أي تقاطعهما $B \cap \overline{B} = \phi$.

(iv) الحادثة المستحيلة (ϕ)

هي الحادثة التي لا تحتوي على نقاط عينة مـن S وتكتب $\{\ \} = \phi$ مثل عدد طلاب جامعة الملك سعود وأعمارهم أقل من عشر سنوات أو عدد النقاط الـتـي تظهر على وجه حجر النرد وعددها 7 نقاط مثلا.

(v) الحادثة المؤكدة (S) .

وهي الحادثة أو الحدث الذي يحتوي على جميع العناصر الممكنة للتجربـة العشوائية مثل $S = \{1,2,3,4,5,6\}$. لتجربة قذف حجر النرد مرة واحدة .

ملاحظة

ويرمـز لعدد نقـاط العينـة لحادثـة B مثلا بالرمز $n(B)$ فيكون عدد نقاط العينـة للحادثـة $B = \{H\}$ عند القـاء قطعة العملة مرة واحـدة فـإن $n(B) = 1$. ولتجربة قذف حجر النرد مرة واحدة $D = \{\text{ظهور عدد زوجي}\}$ أي $D = \{2,4,6\}$ فإن $n(D) = 3$.

مثال (٣,٨) :

اكتب فراغ العينة عند القاء قطعة عملة ثلاث مرات ثم اكتب عناصر الحوادث التالية :

(i) $A = \{\text{ظهور صورة واحدة فقط}\}$

(ii) B = {ظهور صورة على الاكثر}

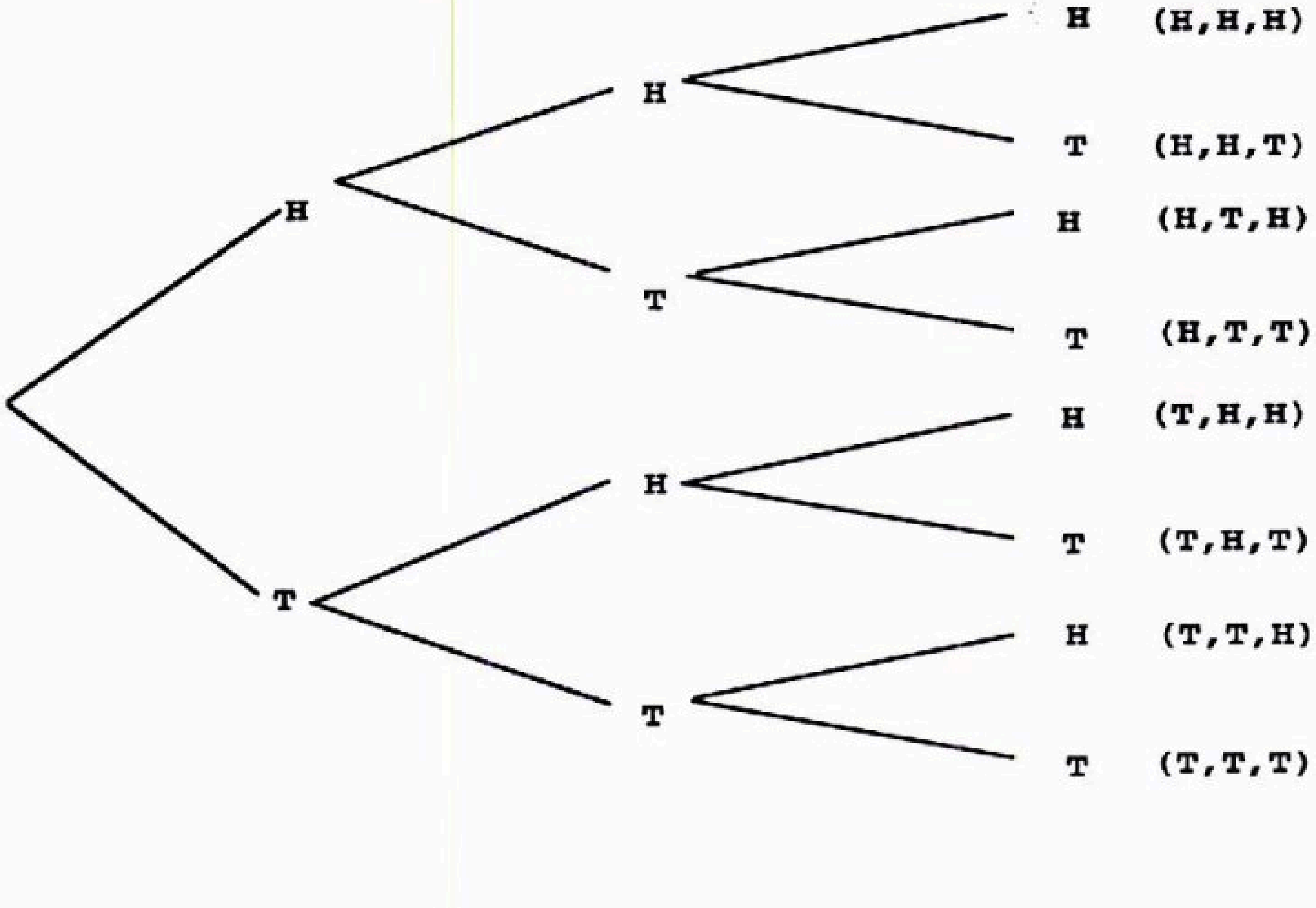
(iii) C = {ظهور صورتين على الأقل}

(iv) $\overline{A},\ \overline{A} \cap B,\ A \cap C,\ A \cup B$

الحل

نستخدم الشجرة البيانية لتوضيح فراغ العينة S كما يلي :

الرمية الثالثة الرمية الثانية الرمية الأولى



نكتب العناصر المرتبة (H,H,H) للسهولة بالرمز HHH وهكذا لباقي العناصر المرتبة فإن فراغ العينة S يكتب كما يلي :

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

فإن عدد عناصر S هو $N(S) = 8$

وإن الحوادث وعدد عناصرها توضح كما يلي :

$$A = \{HTT, THT, TTH\}, \quad n(A) = 3$$

$$B = \{TTT, HTT, THT, TTH\}, \quad n(B) = 4$$

$$C = \{HHT, HTH, THH, HHH\}, \quad n(c) = 4$$

$$\overline{A} = S - A = \{TTT, HHT, THH, HTH, HHH\}, n(\overline{A}) = 5$$

$$\overline{A} \cap B = \{TTT\}, \quad n(\overline{A} \cap B) = 1$$

$$A \cap C = \{\quad\} = \phi \quad , \quad n(A \cap C) = n(\phi) = 0$$

$$A \cup B = \{TTT, HTT, THT, TTH\}, \quad n(A \cup B) = 4$$

مثال (٩ , ٣)

اكتب فراغ العينة وعدد نقاطها وكذلك الحوادث التالية وذلك عند قذف حجر النرد مرتين متتاليتين. حيث الرمية الأولى يرمز لها بالرمز x والرمية الثانية بالرمز y.

$$A = \{(x,y)\ ;\ x + y \leq 2\}$$
$$B = \{(x,y)\ ;\ x = y\}$$
$$C = \{(x,y)\ ;\ x + y = 13\}$$
$$D = \{x,y)\ ;\ y = 3\}$$

الحل

نوجد فراغ العينة S باستخدام شبكة التربيع كما يلي :

الرمية الثانية

6	1,6	2,6	3,6	4,6	5,6	6,6
5	1,5	2,5	3,5	4,5	5,5	6,5
4	1,4	2,4	3,4	4,4	5,4	6,4
3	1,3	2,3	3,3	4,3	5,3	6,3
2	1,2	2,2	3,2	4,2	5,2	6,2
1	1,1	2,1	3,1	4,1	5,1	6,1
	1	2	3	4	5	6

الرمية الأولى

$$S = \{(1,1), (1,2), \ldots, (6,6)\} ; \quad n(S) = 36$$

$$A\{(1,1)\} \quad , \; n(A) = 1$$

$$B = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\} , \; n(B) = 6$$

$$C = \{ \quad \} = \phi , \quad n(\phi) = 0$$

$$D = \{(1,3), (2,3), (3,3), (4,3), (5,3), (6,3)\}, \; n(D) = 6$$

(٣ , ٤ , ٥) الحالات المواتية (Favourable cases)

هي الحالات التي تؤدي إلى تحقق حدث معين مثلا ظهور الصورة للحادثة A = {H} عند رمي قطعة العملة مرة واحدة فهي حالة مواتية وظهور الرقم 2 أو 4 أو 6 عند رمي حجر النرد تعتبر حالات مواتية لحدث ظهور رقم زوجي B = {2,4,6} وهكذا .

(٣ , ٤ , ٦) الحالات المتماثلة أو المتساوية الفرصة (Equaly likely cases)

إذا كان لدينا تجربة عشوائية وكانت جميع نتائج التجربة متساوية الفرصة في الظهور. مثل ظهور الصورة (H) أو الكتابة (T) عند رمي قطعة عملة متزنة فإن فرصة ظهور الصورة تكون مساوية لفرصة ظهور الكتابة ويعتبر ظهور الصورة والكتابة حالتين متماثلتين . وعند رمي حجر نرد متجانس ومنتظم فإن ظهور أي وجه من الأوجه يعتبر حالة متماثلة وهكذا .

(٣ , ٤ , ٧) الحوادث الشاملة (Exhaustive events)

إذا كان لدينا مجموعة حوادث $A_1, A_2, ..., A_n$ عند اجراء تجربة عشوائية ما. فإنه يقال لهذه الحوادث السابقة إنها شاملة إذا كان لا بد من حدوث إحداهم عند إجراء التجربة واتحادهم يساوي فراغ العينة S أي أن :

$$A_1 \cup A_2 \dots \cup A_n = S$$

مثـال ذلك عند القـاء حجر نرد فـإن الحوادث البسيطـة $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$ تعتبر حوادث شاملة .

(٣ , ٤ , ٨) الحوادث المتنافية (المنفصلة) (Mutually exclusive events)

إذا استحال وقوع حادثتين أو أكثر في آن واحد ، فإنه يقال إن هذه الحوادث متنافية . مثل ظهور الصورة (H) والكتابة (T) معا عند رمي قطعة عملة ، أو ظهور وجهين معا عند القاء حجر النرد . ويكون التقاطع لهاتين الحادثتين أو أكثر مساويا ϕ مثال على ذلك عند رمي قطعة العملة فإن $\{H\} \cap \{T\} = \phi$ وهكذا .

(٣ , ٥) التعريف الكلاسيكي للاحتمال

(Classical Definition)

إذا كان لدينا تجربة عشوائية جميع نتائجها متماثلة أي أنها متساوية الفرصة في الظهور، وكان فراغ العينة S وعدد نقاط العينة له محدود ويساوي $n(S)$ نقطة عينة، وكان لدينا الحادثة A من S. فإن احتمال الحادثة A ويرمز لـه $P(A)$ (ويقرأ احتمال A) وكانت $n(A)$ عدد نقاط العينة للحادثة A فإن $P(A)$ تحسب كما يلي:

$$P(A) = \frac{n(A)}{n(S)} \qquad (٣ , ٧)$$

مثال (٣ , ١٠)

إذا ألقيت قطعة عملة متزنة مرة واحدة . أوجد احتمال ظهور الصورة .

الحل

فراغ العينة هو $S = \{H,T\}$ وأن $n(S) = 2$

الحادثة A تمثل ظهور الصورة $A = \{H\}$ وأن $n(A) = 1$ فإن احتمال A ويكتب $P(A)$ هو

$$P(A) = \frac{n(A)}{n(S)} = \frac{1}{2}$$

(٣ , ٦) الاحتمال التجريبي (النسبي)
(Experimental Definition)

من عيب التعريف الكلاسيكي أنه مبني على تساوي الفرص في الظهور لكل نتيجة من جميع نتائج التجربة العشوائية ، وهذا الشرط غير محقق لجميع التجارب العشوائية ، لذلك دعت الحاجة إلى تعريف أعم بحيث يشمل التجارب العشوائية غير متساوية الفرص في الظهور لكل نتيجة لها . وهذا التعريف مبني على أساس إجراء التجربة العشوائية عددا كبيرا من المرات ، بعد ذلك يمكن استنباط قيمة الاحتمال للحادثة أو الحدث A فمثلا ، فإذا اجريت التجربة n من المرات وكان عدد مرات حدوث A يساوي m مرة فإن التكرار النسبي للحدث A هو $\frac{m}{n}$. وبتكرار التجربة عددا كبيرا من المرات فإن التكرار النسبي يستقر عند مقدار ثابت يمثل احتمال الحدث A أي أن

$$P(A) = \lim_{n\to\infty} \frac{n(A)}{n(S)} \qquad (٣ , ٨)$$

فمثلا عند رمي قطعة عملة متزنة عدد n من المرات وكانت الحادثة A (تمثل ظهور الصورة H) فإن احتمال A هو

$$P(A)=\lim_{n\to\infty}\frac{n(A)}{n(S)}=\frac{1}{2}$$

مثال (٣, ١١)

في مصنع للمصابيح الكهربائية وجد من بين كل 100 مصباح يوجد 6 مصابيح محروقة . إذا اخترنا مصباح من إنتاج هذا المصنع ما هو احتمال أن يكون محروقا؟

الحل

نعرف الحدث A بأنه {المصباح محروق} = A

$$P(A)=\lim_{n\to\infty}\frac{n(A)}{n(S)}=\frac{m}{n}=\frac{6}{100}=0.06$$

(٣, ٧) التعريف الحديث للاحتمال (مسلمات الاحتمال)
(Probability Axioms)

من عيب التعريف التجريبي أننا لا نستطيع أن نؤكد أننا سوف نحصل على النسبة نفسها إذا أجريب التجربة العشوائية n من المرات في وقت آخر ، كذلك من صعوباته من وجهة النظر الرياضية ، فقد تكون هذه النهاية غير موجودة بالفعل . لذلك فقد ظهر التعريف الحديث للاحتمال الذي يعتمد على مسلمات تتناسب مع فكرتنا لمعنى الاحتمال . تصاغ هذه المسلمات في ثلاث مسلمات رئيسية نوردها كما يلي :

إذا كانت S فراغ العينة لتجربة عشوائية ما وكانت B و A أي حادثتين من S عندئذ تسمى P دالة احتمال ويسمى العدد P(A) باحتمال الحادثة A إذا تحققت

المسلمات التالية :

(أ) المسلمة الأولى: لأي حادثة A فإن:

$$\text{(I)}\ \ P(A) \geq 0$$

(ب) المسلمة الثانية: لفراغ العينة S فإن:

$$\text{(II)}\ \ P(S) = 1$$

(ج) المسلمة الثالثة: لأي حادثتين متنافيتين B و A حيث $A \cap B = \phi$ فإن :

$$\text{(III)}\ P(A \cup B) = P(A) + P(B)$$

ملاحظة

ويمكن تعميم المسلمة الثالثة لأي عدد من الحوادث المتنافية مثنى مثنى كما يلي:

لأي تجزيء من S من الحوادث المتنافية $A_1, A_2, \ldots..$ حيث $A_i \cap A_j = \phi$ لأي $i \neq j$ فإن

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i),\ \ i = 1, 2, \cdots$$

ونستخدم مسلمات الاحتمال السابقة في اثبات بعض النظريات الاحتمالية كما يلي:

نظرية (١)

احتمال حدوث الحادثة المستحيلة ϕ يساوي صفرا أي أن :

$$P(\phi) = 0 \qquad (٣,٩)$$

البرهان

الحادثتان الأكيدة S والمستحيلة ϕ تحقق المسلمة (III)

حيث $S \cup \phi = S \;;\quad S \cap \phi = \phi$

فإن $P(S \cup \phi) = P(S) + P(\phi)$

أي أن

$$P(S) = P(S) + P(\phi)$$

أي أن

$$P(\phi) = P(S) - P(S) = 0$$

وهو المطلوب .

نظرية (٢)

احتمال حدوث الحدث A مضافا إليه احتمال حدوث الحـدث $\overline{A}$ يساوي الواحد الصحيح أي أن :

$$P(A) + P(\overline{A}) = 1 \qquad (٣,١٠)$$

البرهان

الحدثان A و $\overline{A}$ تحققا المسلمة (III)

حيث $A \cap \overline{A} = \phi$ وأن

$$A \cup \overline{A} = S \qquad (٣,١١)$$

بإدخال دالة الاحتمال P على طرفي (٣,١١) واستخدام المسلمة (II) و المسلمة (III)

فإن $P(A \cup \overline{A}) = P(S)$

$$P(A) + P(\overline{A}) = 1$$

وهو المطلوب .

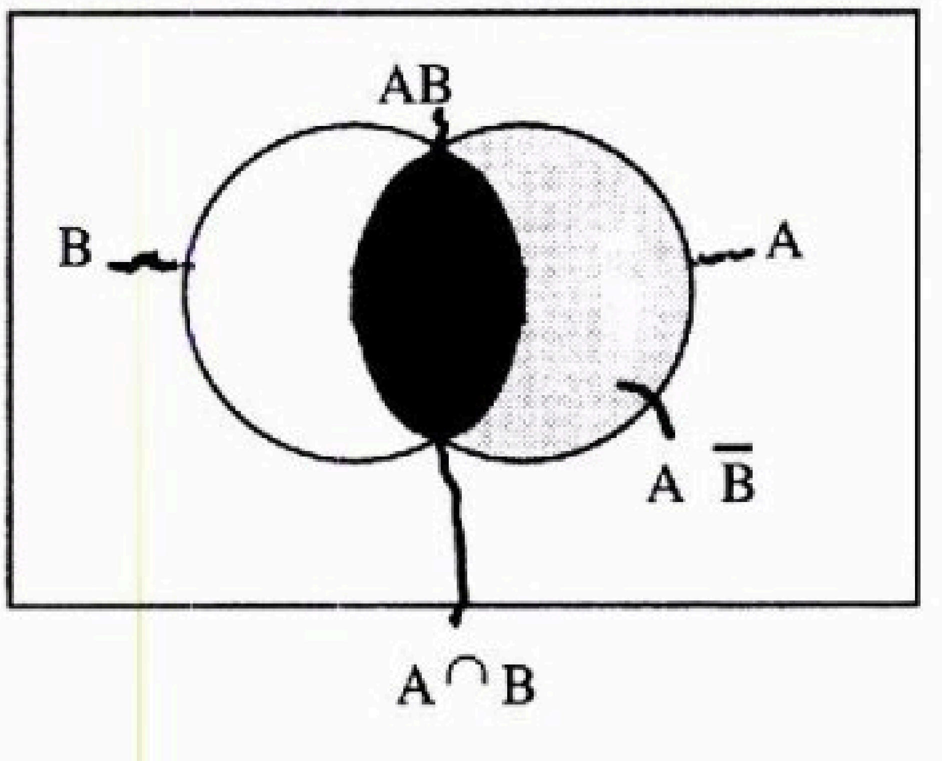
نظرية (٣).

لأي حدثين B و A ، فإن حدوث الحدث A وعدم حدوث الحدث B يساوي احتمال الحدث A مطروحا منه احتمال الحدثين A, B معا أي أن

$$P(A \cap \overline{B}) = P(A) - P(A \cap B) \tag{٣, ١٢}$$

البرهان

نوضح الحدثين A,B بشكل فن (٣, ٤) كما يلي :



شكل (٣, ٤)

ستختصر فيما يلي التقاطع $A \cap B$ بالمقدار AB والتقاطع $A \cap \overline{B}$ بالمقدار $A\,\overline{B}$ وكذلك لأي تقاطع آخر .

فمن شكل فن نجد أن الحدثين $A\,\overline{B}$ ، AB متنافيين حيث $A\,\overline{B} \cap A\,B = \phi$ فإن

$$A\,\overline{B} \cup A\,B = A \tag{٣, ١٣}$$

بادخال دالة الاحتمال P على طرفي (٣,١٣) واستخدام المسلمة (III) فإن

$$P(A\overline{B}) + P(AB) = P(A)$$

أي أن

$$P(A\overline{B}) = P(A) - P(AB)$$

وهو المطلوب .

مثال (٣,١٢)

إذا كان احتمال نجاح محمد هو $\frac{3}{5}$ واحتمال نجاح احمد ومحمد معا هو $\frac{1}{5}$ اوجد احتمال نجاح محمد ورسوب احمد .

الحل

نعرف الأحداث التالية :

$P(A) = \frac{3}{5}$ و (نجاح محمد) =A

$\overline{B}$ = (رسوب أحمد) و (نجاح أحمد) = B

$P(AB) = \frac{1}{5}$ و {نجاح محمد وأحمد معا} = AB

{نجاح محمد ورسوب أحمد} = $A\overline{B}$

فإن

$$P(A\overline{B}) = P(A) - P(AB)$$

$$= \frac{3}{5} - \frac{1}{5} = \frac{2}{5}$$

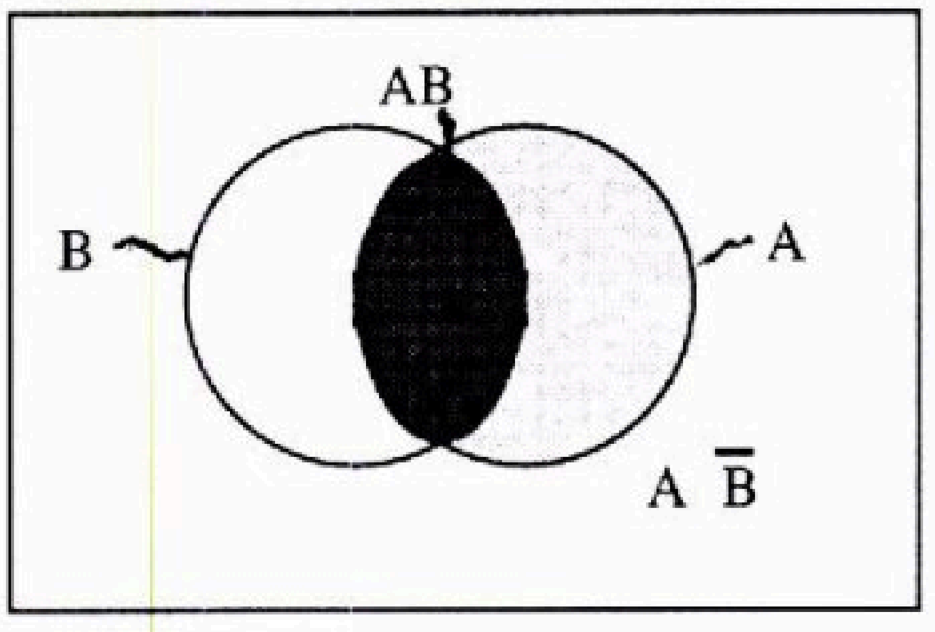
نظرية (٤)

لأي حدثين A, B ، فإن احتمال حدوث أحدهما على الأقل يساوي احتمال A مضافا إليه احتمال B مطروحا منهما احتمال A,B معا أي أن

$$P(A \cup B) = P(A) + P(A) - P(AB) \qquad (٣,١٤)$$

البرهان

نوضح الحدثين A, B بشكل فن (٣,٥) كما يلي :



شكل (٣,٥)

اتحاد $A \cup B$ يمكن تقسيمه إلى حدثين متنافيين $B, A\overline{B}$ كما هو موضح بشكل فن (٣,٥) فإن

$$A \cup B = B \cup A\overline{B} \qquad (٣,١٥)$$

بادخال دالة الاحتمال P على طرفي (٣,١٥) واستخدام المسلمة (III) فنحصل على:

$$P(A \cup B) = P(B) + P(A\overline{B}) \qquad (٣,١٦)$$

بالتعويض بقيمة $P(A\overline{B})$ من العلاقة (٣,١٢) نحصل على

$$P(A \cup B) = P(B) + P(A) - P(AB)$$

وهو المطلوب .

مثال (٣,١٣)

إذا كان احتمال نجاح محمد في مقرر للإحصـاء هـو $\frac{3}{8}$ واحتمال نجـاح محمد وأحمد في نفس المقرر $\frac{2}{8}$ واحتمال نجاح أحدهما على الأقل هـو $\frac{5}{8}$. فأوجد احتمال نجاح أحمد في هذا المقرر .

الحل

نكتب الأحداث التالية : $P(A) = \frac{3}{8}$ و {نجاح محمد} = A

$P(B) = ?$ و {نجاح أحمد} = B

$P(AB) = \frac{2}{8}$ و {نجاح محمد وأحمد معا} = AB

$P(A \cup B) = \frac{5}{8}$ و {نجاح محمد أو احمد} = $A \cup B$

فإن

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

$$\frac{5}{8} = \frac{3}{8} + P(B) - \frac{2}{8}$$

وبقليل من الاختصار نجد أن :

$$P(B) = \frac{5}{8} + \frac{2}{8} - \frac{3}{8} = \frac{1}{2}$$

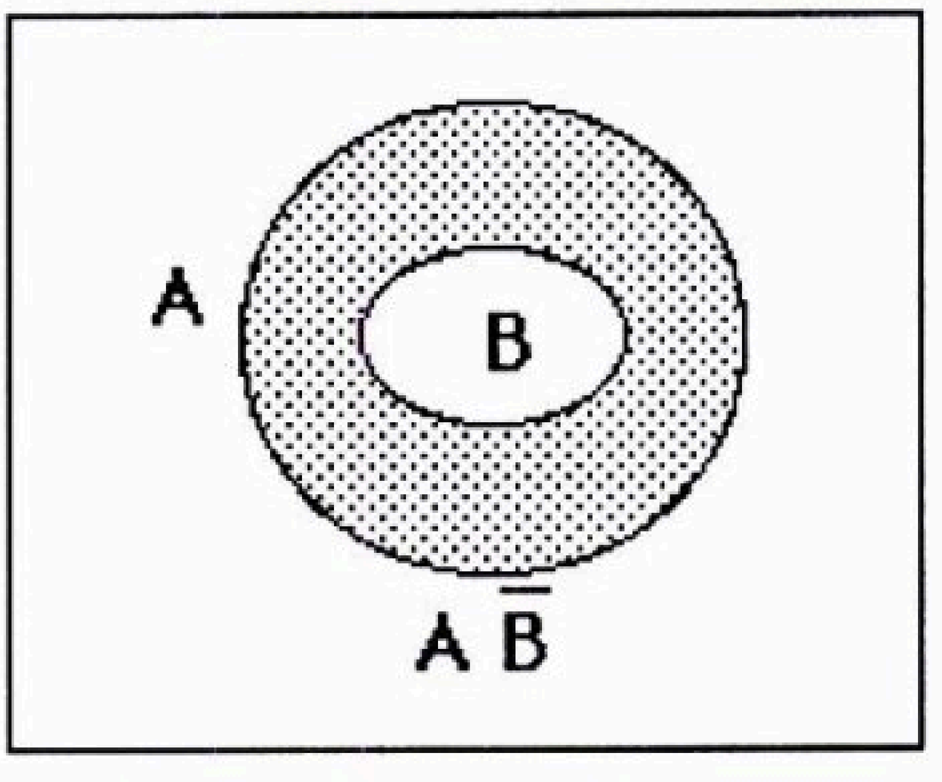
نظرية (٥)

إذا كان الحدث B مجموعة جزئية من الحدث A فإن

$$P(B) \leq P(A) \qquad (٣ , ١٧)$$

البرهان

نوضح الحدثين A, B في شكل فن (٣ , ٦) كما يلي :



شكل (٣ , ٦)

من شكل (٣ , ٦) واضح أن الحــدث A هو عبارة عن اتحاد الحدثين المتــنــافــيــين $B, A\overline{B}$ حيث $B \cap A\overline{B} = \phi$ فإن

$$A = B \cup A\overline{B} \qquad (٣ , ١٨)$$

بإدخال دالة الاحتمال P على العلاقة (٣ , ١٨) واستخدام المسلمـة (III) لنحصل على

$$P(A) = P(B) + P(A\overline{B})$$

أي أن

$$P(A) - P(B) = P(A\overline{B}) \qquad (٣ , ١٩)$$

من المسلمة (I) في العلاقة (٣ , ١٩) فإن $P(A\overline{B}) \geq 0$ ينتج أن

$$P(A) - P(B) \geq 0$$

أي أن

$$P(B) \leq P(A)$$

وهو المطلوب .

مثال (٣ , ١٤)

إذا كان لدينا حدثين B و A بحيث كان

$$P(A) = 0.2,\ \ P(B) = 0.4,\ \ P(A \cup B) = 0.5$$

أوجد

$$\text{(i)}\ \ P(AB), \qquad \text{(ii)}\ P(\overline{A}\,\overline{B}), \qquad \text{(iii)}\ P(A\overline{B})$$

الحل

$$\text{(i)}\ \ P(AB) = P(A) + P(B) - P(A \cup B)$$

$$= 0.2 + 0.4 - 0.5 = 0.1$$

$$\text{(ii) } P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B)$$
$$= 1 - 0.5 = 0.5$$

$$\text{(iii) } P(A\overline{B}) = P(A) - P(AB)$$
$$= 0.2 - 0.1 = 0.1$$

(٣,٨) الاحتمال الشرطي والاستقلال
(Conditional Probability and Independent Events)

(٣,٨,١) الاحتمال الشرطي

يعرف عندما يكون فراغ العينة S محدود . ويحدث في الحياة العملية أننا نحتاج لإيجاد احتمال الحدث A مثلا علما بوقوع الحدث B . ومن الأمثلة على ذلك عند سحب كرتين بدون إرجاع من صندوق به n كـــرة ذات لونين الأبيـــض (n_1 كـــرة) والأسود (n_2 كـــرة) . المطلوب ايجاد احتمال أن تكون الكـــرة الثانية بيضاء وعلما بأن الكرة الأولى كانت بيضاء أيضا . وأحيانا يكون مطلوب حساب احتمال مصباح أن يستمر صالحا لمدة 150 ساعة علما بأنه ظل صالحا لمدة 100 ساعة .

مما سبق يكون الاحتمال للحدث A منسوبا لفراغ عينة جديد B حيث الحدث B هو $B \subset S$ ونوضح ذلك بمثال عددي كما يلي :

مثال (٣,١٥)

معرض سيارات به نوعين من السيارات كل نوع يوجد به عدد من السيارات

الجديدة والمستعملة لمدة عام كما هو موضح بالجدول (٣,١) كما يلي :

النـــوع	النوع الأول	النوع الثاني	المجموع
سيارات جديدة	30	35	65
مستعملة لمدة عام	10	25	35
المجموع	40	60	100

سحبنا سيارة بطريقة عشوائية فإننا نلاحظ أن :

(i) احتمال أن تكون السيارة من النوع الأول وجديدة = $\frac{30}{100}$.

(ii) فإذا علمنا أن السيارة من النوع الأول فإن احتمال أن تكون جديدة يساوي = $\frac{30}{40}$.

والاحتمال (i) ناتج من أن عدد الحالات المواتيــة 30 ، وعدد الحالات الممكنة هو 100، والاحتمال (ii) هو احتمال أن تكون السيارة جديدة إذا علم أنها من النــوع الأول ويسمى احتمالا شرطيا . فإذا رمزنا للحدث A بأن السيارة المختارة جديدة، والحدث B بأن السيارة المختارة من النوع الأول ، فإن احتمال (i) هو P(AB) وأن الاحتمال (ii) هو P(A/B) ويقرأ إحتمال حدوث A بشرط حدوث B ونلاحظ أن :

$$P(A/B) = \frac{30}{40} = (\frac{30}{100}) / (\frac{40}{100}) = \frac{P(AB)}{P(B)}$$

وبذلك يكون تعريف احتمال حدوث A بشرط وقوع الحدث B هو P(A/B) أي أن:

$$P(A/B) = \frac{P(AB)}{P(B)} \qquad (٣,١٩)$$

بحيث $P(B) \neq \phi$.

مثال (٣,١٦)

الحدثين A , B بحيث

$$P(\overline{B}) = \frac{5}{8}, P(A \cup B) = \frac{3}{4},\ P(A) = \frac{1}{2}$$

احسب قيمة كل من :

$$P(B),\ \ P(A/B),\ \ P(B/\ \overline{B})$$

الحل

$$P(B) = 1 - P(\overline{B}) = 1 - \frac{5}{8} = \frac{3}{8}$$

$$P(A/B) = \frac{P(AB)}{P(B)} = \frac{P(A) + P(B) - P(A \cup B)}{3/8}$$

$$= \frac{\frac{1}{2} + \frac{3}{8} - \frac{3}{4}}{3/8} = \frac{1}{3}$$

$$P(B/\overline{A}) = \frac{P(B\overline{A})}{P(\overline{A})} = \frac{P(B) - P(AB)}{1 - P(A)}$$

$$= \frac{3/8 - 1/8}{1/2} = \frac{1}{2}$$

(٣,٨,٢) احتمال الأحداث المستقلة (Probability of independent events)

يقال للحدثين A, B مستقلان إذا كان حدوث أحدهما لا يؤثر في حدوث الآخر . فمثلا عند رمي قطعة العملة مرتين فإن ظهور الصورة في الرمية الثانية لا يتأثر بالناتج في الرمية الأولى . ونعبر عن استقلال A, B كما يلي :

$$P(B/A) = P(B) \quad (٣,٢٠)$$

وتصبح العلاقة (٣,١٩) للاحتمال الشرطي تحت شرط الاستقلال (٣,٢٠) كما يلي :

$$P(A) = \frac{P(AB)}{P(B)}$$

أي أن

$$P(AB) = P(A)\,P(B) \quad (٣,٢١)$$

مثال (٣,١٧)

احتمال أن يصيب محمد الهدف هو $\frac{3}{5}$ واحتمال أن يصيب ابراهيم الهدف $\frac{1}{5}$. فأوجد احتمالات الأحداث التالية :

ا) أن يصيبا الهدف معًا

ب) على الأقل يصيب الهدف أحدهما .

جـ) محمد يصيب الهدف وابراهيم لا يصيب

الحل

إصابة محمد الهدف مستقلة عن إصابة ابراهيم الهدف نكتب الحوادث التالية :

A = {محمد يصيب الهدف} و B = {ابراهيم يصيب الهدف}

AB = {محمد وابراهيم يصيبا معا} و $A \cup B$ = {على الأقل أحدهما يصيب} .

ا) $P(AB) = P(A) \,.\, P(B) = \frac{3}{5} \,.\, \frac{1}{5} = 0.12$

ب) $$P(A \cup B) = P(A) + P(B) = P(AB)$$
$$= 0.6 + 0.2 - 0.12= 0.68$$

جـ) $$P(A\overline{B}) = P(B) - P(AB)$$
$$= 0.2 - 0.12= 0.08$$

مثال (٣, ١٨)

صندوق وجد به 15 كرة بيضاء و 25 كرة سوداء أخذت كرتان واحدة بعد الأخرى، أوجد في حالتي السحب بدون إرجاع والسحب بإرجاع ما يلي :

أ) احتمال أن تكون الكرتان لونهما أبيض .

ب) احتمال أن تكون الكرتان من نفس اللون .

جـ) احتمال أن تكون الكرة الأولى بيضاء والثانية سوداء .

الحل

نكتب رموز للأحداث التالية :

$W_1 =$ {الكرة الأولى بيضاء} و $W_2 =$ {الكرة الثانية بيضاء}

$B_1 =$ {الكرة الأولى سوداء} و $B_2 =$ {الكرة الثانية سوداء}

$W_1W_2 =$ {الكرتان لونهما أبيض}

$W_1W_2 \cup B_1B_2 =$ {الكرتان من نفس اللون}

$W_1B_2 =$ {الأولى بيضاء والثانية سوداء}

أولا : في حالة السحب بدون إرجاع (*الحوادث غير مستقلة*) نجد أن :

أ) $$P(W_1W_2) = P(W_1)\,P(W_2/W_1)$$
$$= \frac{15}{40} \cdot \frac{14}{39}$$

ب) $P(W_1W_2 \cup B_1B_2) = P(W_1W_2) + P(B_1B_2)$

$$= \frac{15}{40} \cdot \frac{14}{39} + \frac{25}{40} \cdot \frac{24}{39}$$

جـ) $P(W_1B_2) = P(W_1)P(B_2 / W_1)$

$$= \frac{15}{40} \cdot \frac{25}{39}$$

ثانيا: في حالة السحب بإرجاع (*الحوادث مستقلة*) نجد أن :

أ) $P(W_1W_2) = P(W_1)\,P(W_2)$

$$= \frac{15}{40} \cdot \frac{15}{40} = \left(\frac{15}{40}\right)^2$$

ب) $P(W_1W_2 \cup B_1B_2) = \left(\frac{15}{40}\right)^2 + \left(\frac{25}{40}\right)^2$

جـ) $P(W_1B_2) = \frac{15}{40} \cdot \frac{25}{40}$

(٩ , ٣) الاحتمال الكلي

Total Probability

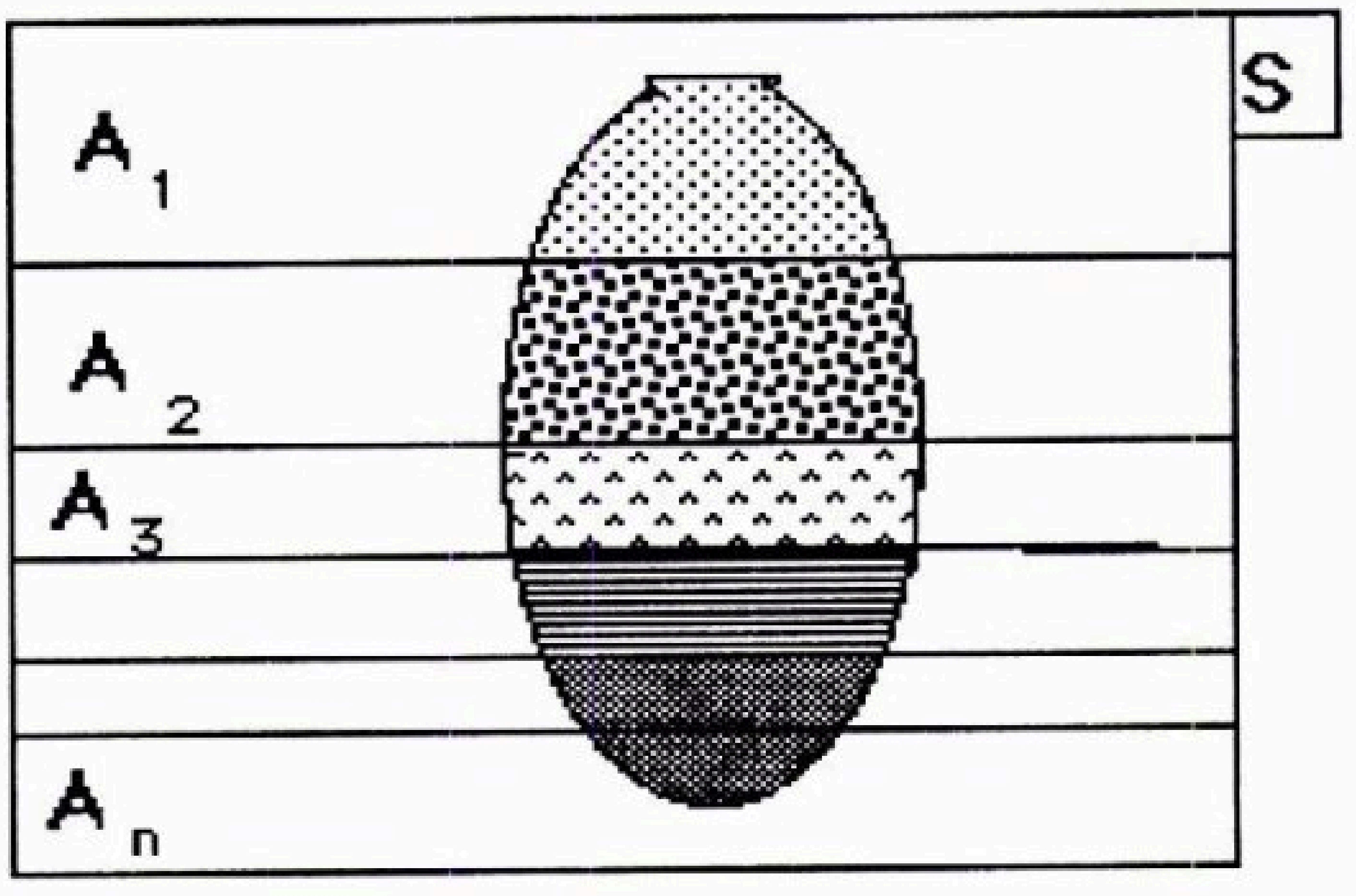
لنفرض أن الحوادث غير الخالية $A_1, A_2, ..., A_n$ تشكل تجزيء لفضاء العينة S أي متنافية (منفصلة) حيث إن :

$$S = A_1 \cup A_2 \cup ... \cup A_n, \quad A_i \cap A_j = \phi, \quad i \neq j.$$

فإنه يمكن التعبير عن أي حادثة B معرفة على نفس فراغ العينة S بدلالة تقاطعات هذه الحادثة مع كل من حوادث التجزيء كما يلي :

$$B = B \cap S = B \cap (A_1 \cup A_2 \dots \cup A_n)$$
$$= BA_1 \cup BA_2 \dots BA_n$$

كما يوضح ذلك بالجزء المظلل في شكل (٣,٧) كما يلي :



شكل (٣,٧)

ووفقا للمسلمة الثالثة نجد أن

$$P(B) = P(BA_1 \cup BA_2 \dots \cup BA_n)$$
$$= P(BA_1) + P(BA_2) + \dots + P(BA_n)$$

وباستخدام الاحتمال الشرطي العلاقة (٣,١٩) نجد أن

$$P(B) = P(B/A_1)\,P(A_1) + P(B/A_2)\,P(A_2) + \dots + P(B/A_n)\,P(A_n)$$

$$P(B) = \sum_{i=1}^{n} P(B/A_i)\,P(A_i) \qquad (٣,٢٢)$$

والعلاقة (٣,٢٢) السابقة تسمى قانون الاحتمال الكلي .

مثال (١٩ , ٣)

أربع طرق تؤدي إلى بئر ماء . إختار شخص أحد هذه الطرق عشوائيا . فإذا إختار الطريق الأول A فإن احتمال وصوله إلى البئر يساوي $\frac{1}{8}$. وإذا إختار الطريق الثاني B فإن احتمال وصوله يساوي $\frac{1}{6}$. أما إذا إختار الطريق الثالث C فإن احتمال وصوله يساوي $\frac{1}{4}$ وأخيرا الرابع D احتمال وصوله يساوي $\frac{9}{10}$.

أوجد احتمال أن يصل الشخص إلى بئر الماء .

الحل

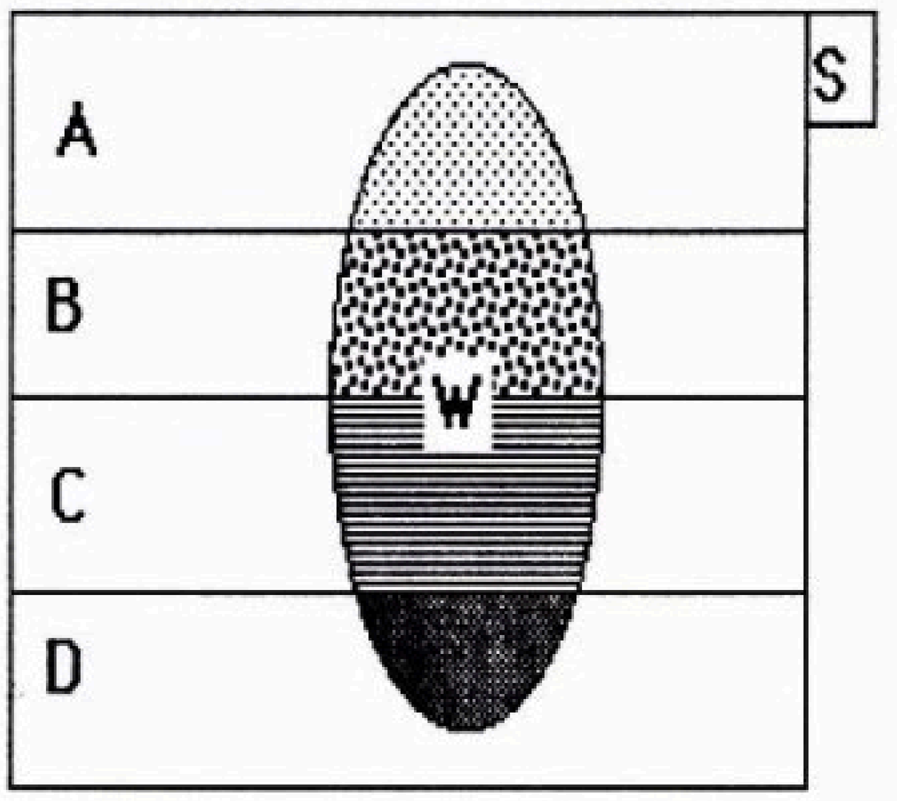
نرمز للحدث A هو اختيار الطريق الأول .

ونرمز للحدث B هو اختيار الطريق الثاني .

ونرمز للحدث C هو اختيار الطريق الثالث .

ونرمز للحدث D هو اختيار الطريق الرابع .

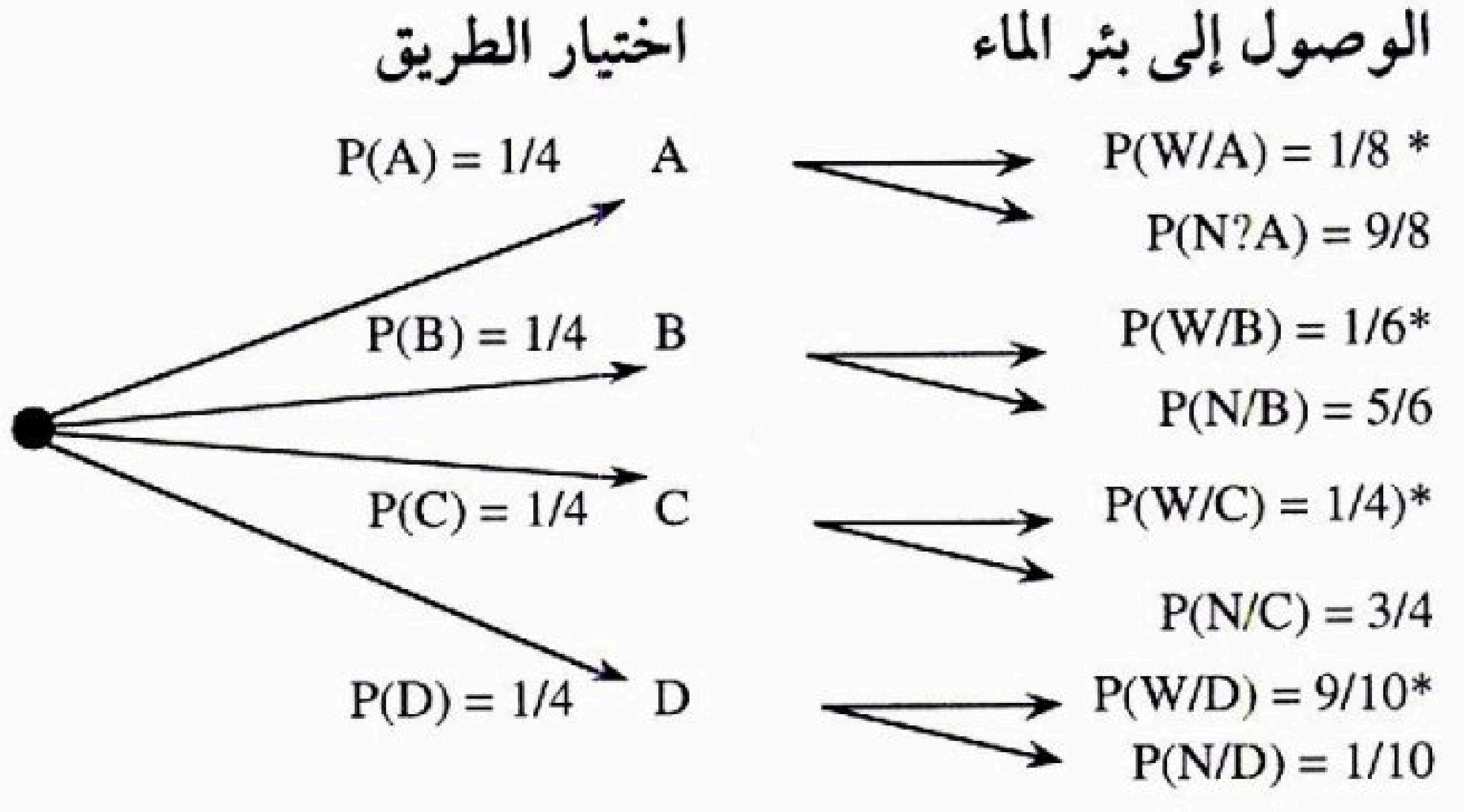
ونلاحظ أن الأحداث A, B, C, D تشكل تجزيء لفضاء العينة S المرافق لتجربة الاختيار العشوائي للوصول إلى بئر الماء وليكن الحدث W يمثل الوصول إلى بئر الماء ويبين الحدث W بالجزء المظلل في الشكل التالي :



ويحسب الاحتمال الكلي P(W) بالاستعانة بشكل فن السابق والعلاقـة (٣,٢٢) كما يلي :

$$P(W) = P(AW) + P(BW) + P(CW) + P(DW)$$
$$= P(A)P(W/A) + P(B)P(W/B) + P(C)P(W/C) + P(D)P(W/D)$$

لتوضيح كيفية حساب الاحتمالات السابقة نرسم الشجرة البيانية التالية :



كما هو موضح بالشجرة البيانية فإن الاحتمال الكلي P(W) هو

$$P(W) = \frac{1}{4}\,(\frac{1}{8}) + \frac{1}{4}\,(\frac{1}{6}) + \frac{1}{4}\,(\frac{1}{4}) + \frac{1}{4}\,(\frac{9}{19})$$

$$= 0.0313 + 0.0417 + 0.0625 + 0.225 = 0.36$$

(٣,١٠) نظرية بيز (Bayes Theorem)

تهتم نظرية بيز حساب احتمال أن يكون هناك سببا ما هو مصدر حـدوث حدث معين نعلم مسبقا بحدوثه . وأن حدوثه يرجع الى عـدد مـن الأسـبـاب المعروف احتمال حدوث كل منها . كما نعلم أيضا احتمال حدوث حدث إذا تحقق سبب ما من هذه الأسباب . فإذا كان لدينا تجزيء S عبارة عن الأحداث حيث يكون $A_1, A_2, \ldots, A_n$

$$S = \bigcup_{i=1}^{n} A_i \quad ، \quad A_i \cap A_j = \phi \quad , \quad i \neq j=1,2,...,n$$

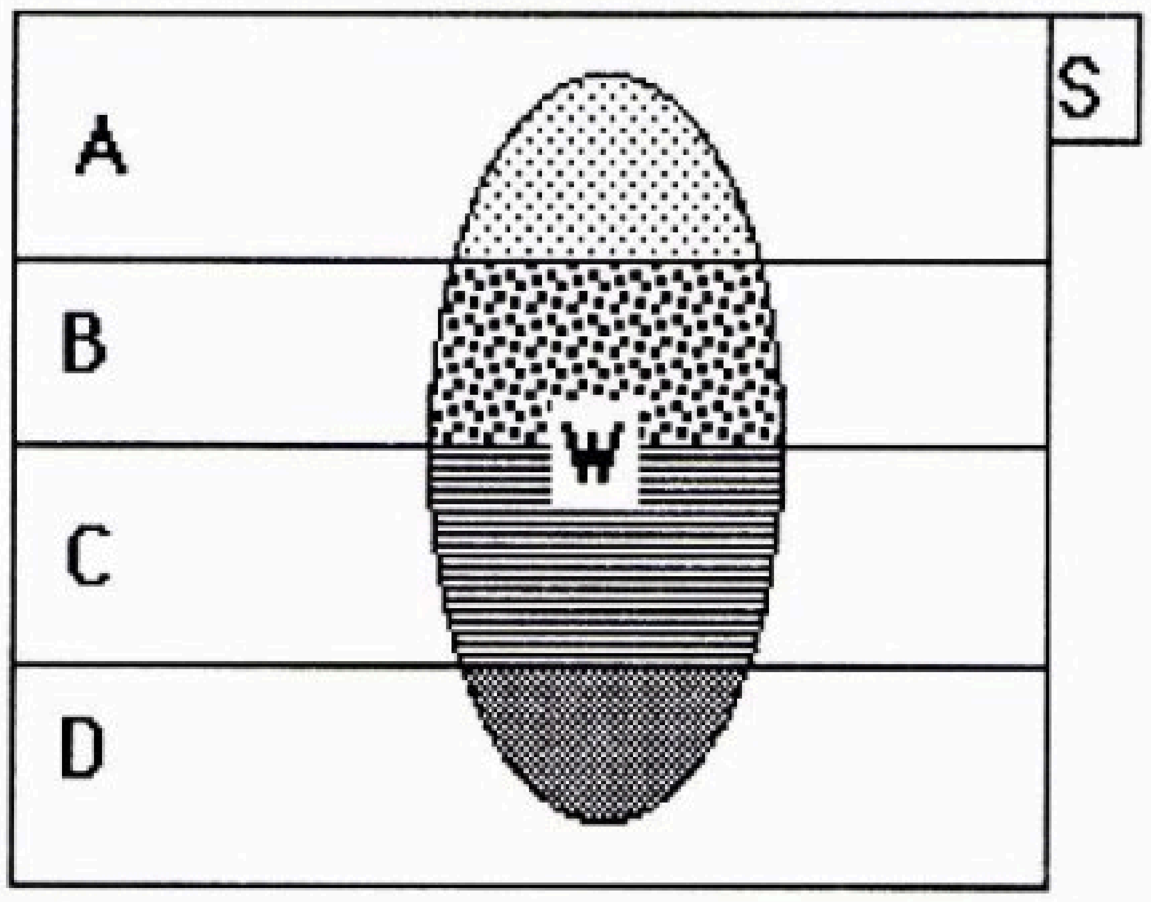
وأن $P(A_i) \neq 0 \quad , \quad i = 1, 2 ,..., n$

لأي حادثة B معرفة على فراغ العينة S بحيث إن $P(B) \neq 0$ فإن

$$P(A_i/B) = \frac{P(B/A_i)\, P(A_i)}{\sum_{i=1}^{n} P(B/A_j)\, P(A_j)} \tag{٣,٢٣}$$

البرهان

نستعين بشكل فن (٣,٨) والتجزيء S هو الأحداث $A_1, A_2 ,..., A_n$ مع الحدث B كما يلي:



شكل (٣,٨)

يمكن التعبير عن الحدث B كما يلي :

$$B = BA_1 \cup BA_2 \ldots \cup BA_i \ldots \cup BA_n$$

$$P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(B/A_i)P(A_i) \qquad (٣,٢٤)$$

ومن تعريف الاحتمال الشرطي (٣,١٩) وبالتعويض بالعلاقة (٣,٢٤) نحصل على:

$$P(A_i/B) = \frac{P(BA_i)}{P(B)} = \frac{P(B/A_i)\,P(A_i)}{\sum_{j=1}^{n} P(B/A_j)\,P(A_j)}$$

وهو المطلوب

مثال (٣,٢٠)

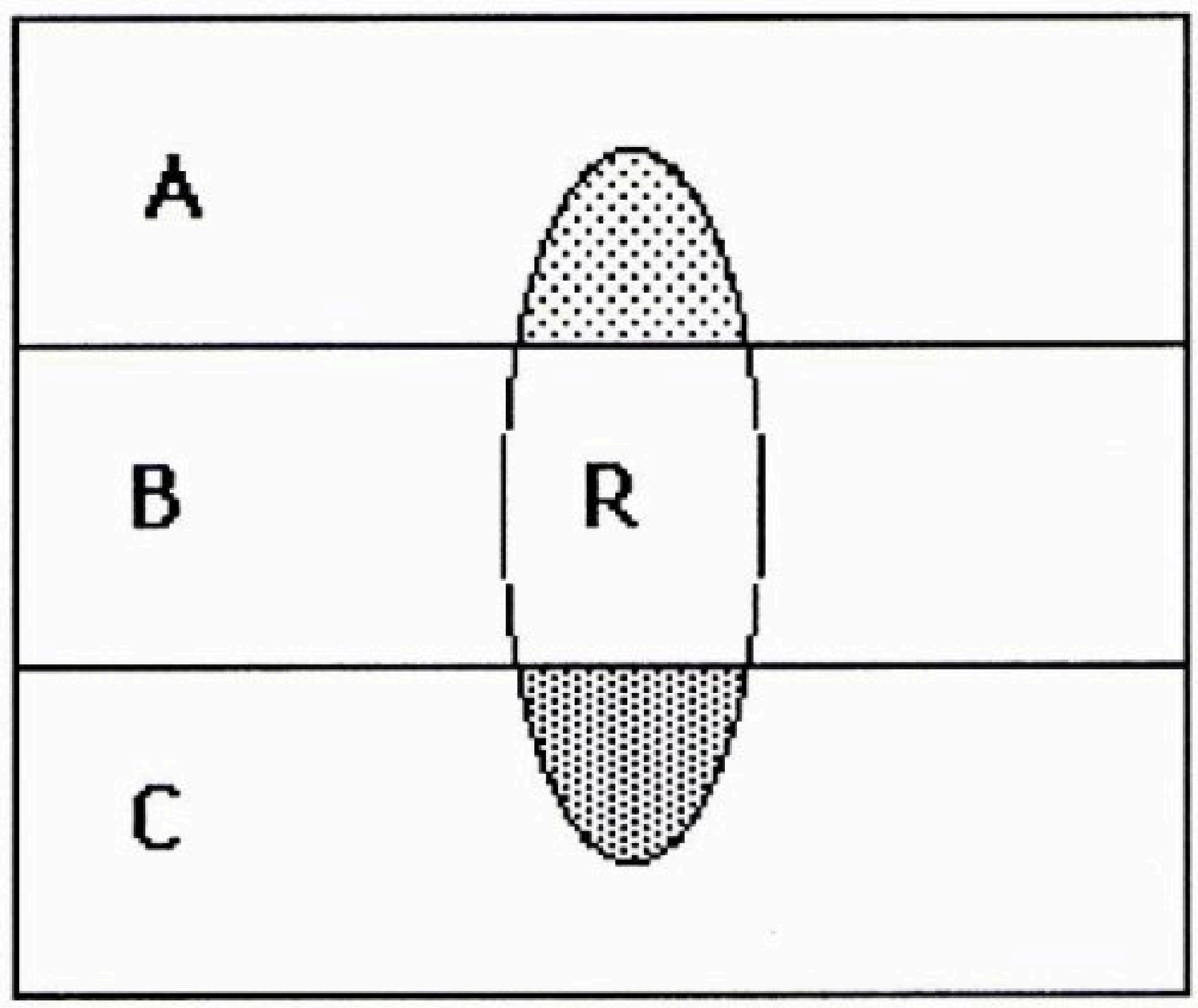
تنوي أسرة قضاء اجازة نهاية الأسبوع في أحدِ الأماكن السياحية A أو B أو C . إذا كان احتمال سقوط المطر في A هو 0.6 وفي B هو 0.7 وفي C هو 0.5 وإذا اختارت الأسرة مكان الاجازة عشوائيا فاحسب :

أ - احتمال أن تقضي الأسرة اجازة ممطرة .

ب - إذا علمت أن الأسرة قضت اجازة ممطرة فما هو احتمال أن تكون قضت اجازتها في المكان B ؟

الحل

نوضح بشكل فن الحوادث التالية :

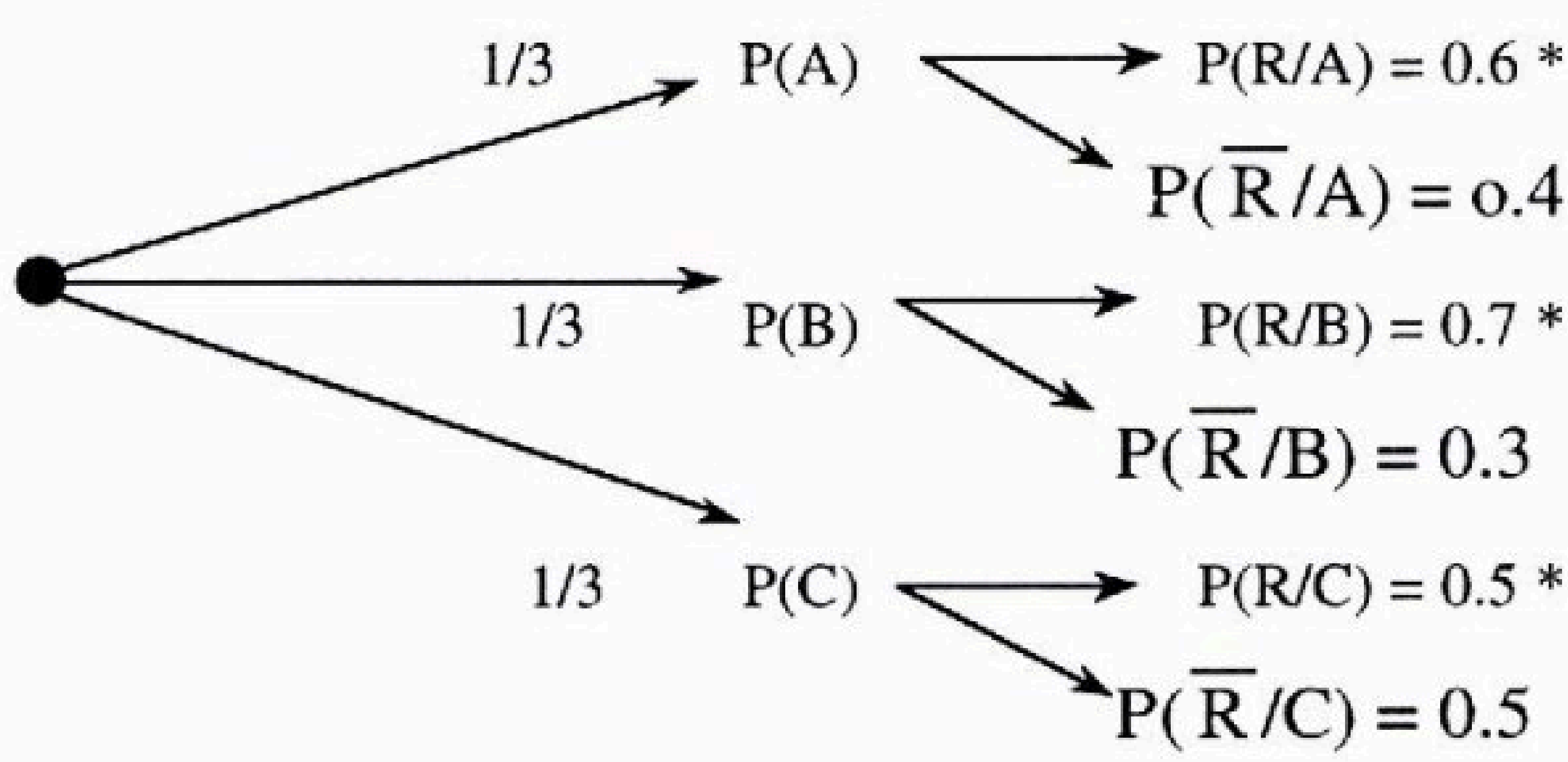


$B = \{$ اختيار المكان B $\}$ $A = \{$ اختيار الأسرة المكان A $\}$

$R = \{$ الاجازة ممطرة $\}$ $C = \{$ اختيار المكان C $\}$

$$P(A) = P(B) = P(C) = \frac{1}{3}$$

ونستخدم الشجرة البيانية لتوضح الاحتمالات التالية :



ا - من العلاقة (٣,٢٤) نجد أن

$$P(R) = \frac{1}{3}(0.6) + \frac{1}{3}(0.7) + \frac{1}{3}(0.5) = \frac{1}{3}(0.18) = 0.06$$

ب - من العلاقة (٣,٢٣) نجد أن

$$P(B/R) = \frac{P(R/B)\,P(B)}{P(R)} = \frac{1/3\,(0.7)}{0.06} = \frac{0.7}{0.18} = 0.39$$

(٣,١١) تمارين

١ - يوجد بين المدينتين B , C ثلاثة طرق وبين المدينتين C , D أربعة طرق بكم طريقة يمكن لشخص قام من B ليصل إلى D ؟

٢ - إذا علم أن أرقام الهواتف الداخلية في جامعة الملك سعود مؤلفة من خمسة أرقام تبدأ دائما بالرقم 7 فما هو عدد الهواتف التي يمكن تركيبها في الحالتين:

ا) التكرار ممكن ب) التكرار غير مسموح .

٣ - لدينا ثمانية طلاب منهم خمسة سعوديين وثلاثة غير سعوديين ، نريد اختيار وفد من أربعة طلاب:

ا) ما هو عدد الطرق الممكنة لاختيار الوفد .

ب) ما هو الطرق الممكنة لاختيار الوفد إذا كان منهم إثنان غير سعوديين.

جـ) ما هو عدد الطرق الممكنة لاختيار الوفد بحيث يكون إثنان على الأكثر غير سعوديين .

٤ - أثبت أن $$\sum_{x=0}^{n} \binom{n}{x} = 2^n$$

٥ - فراغ عينة مكون من الأعداد واحد إلى عشرة . وكانت

$$A = \{2,3,4,\} \ , \ B = \{3,4,5\} \ , \ C = \{5,6,7\}$$

أوجد عناصر الحوادث التالية :

$$(A \cap B) \ , \ (\overline{A} \cup B) \ , \ (\overline{A} \cap \overline{B}) \ , \ A \cap (B \cup C)$$

٦ - اذكر كلا من التعريف التقليدي والتعريف التجريبي للاحتمال وبين الفرق بينهما .

٧ - عرف كل من التجربة العشوائية وفراغ العينة والحدث .

٨ - ا) ما هو الفرق بين الأحداث المتنافية وغير التنافية ؟
ب) ما هو الفرق بين الأحداث المستقلة وغيرالمستقلة ؟

٩ - قطعة نقود متحيزة بحيث فرصة ظهور الصورة ثلاثة أمثال فرصة ظـهـور الكتابة القيت هذه القطعة مرتين :
ا) اكتب فراغ العينة .
ب) أوجد احتمال كل نقطة من نقاط فراغ العينة .
جـ) أوجد احتمال ظهور صورة على الأكثر .
د) أوجد احتمال ظهور صورة على الأقل .

١٠ - إذا كان لدينا حادثتان A , B بحيث كان

$$P(A) = C\ ,\ \ P(A \cup B) = \frac{1}{2}\ ,\ \ P(B) = \frac{1}{3}$$

أوجد ما يأتي:

ا) قيمة C إذا كانت A , B متنافيتين .

ب) قيمة C إذا كانت A , B مستقلتين .

١١ - إذا كان $P(\overline{B}) = 0.7\ ,\ P(A\,\overline{B}) = 0.2$ فأوجد $P(A \cup B)$.

١٢ - إذا علم أن احتمال أن يكون الجو ملبدا بالغيوم هو 0.3 واحتمال أن يكون الجو إما ملبد بالغيوم أو عاصفا هو 0.58 . أوجد الاحتمالات التالية إذا كان احتمال أن الجو عاصف هو 0.4 .

أ) أن يكون الجو ملبدا بالغيوم وعاصفا .

ب) أن يكون الجو ملبدا بالغيوم وغير عاصف .

ج) أن يكون الجو غير ملبد بالغيوم وغير عاصف .

١٣ - G و H حادثان متنافيان وكان $P(H) = 0.3\ ,\ P(G) = 0.4$.

اوجد الاحتمالات $P(H \cup G),\ P(\overline{H} \cup \overline{G}),\ P(\overline{H} \cap G)$

١٤ - صندوق به 8 كرات حمراء و 7 كرات بيضاء . اختيرت منه كرتان . احسب احتمال أن تكون الكرتان المختارتان

ا) لونهما أحمر ب) من اللون نفسه .

١٥ - إذا كان A , B حدثين مستقلين وكان $P(A) = 0.3,\ P(A \cup B) = 0.5$

فأوجد $P(B)$.

١٦ - ثلاثة صناديق تحتوي على عناصر تالفة وغير تالفة كما يلي:

	الصندوق I	الصندوق II	الصندوق III
عدد التالف	7	5	3
عدد السليم	23	15	7

سحب صندوق بطريقة عشوائية وسحب عنصر من ذلك الصندوق أوجد:

ا) احتمال أن العنصر تالفا .

ب) إذا علم أن العنصر تالفا أوجد احتمال أنه من الصندوق II .

١٧ - في إحدى مباني إسكان الجامعة يوجد 150 طالبا منهم 110 يجيدون الإنجليزية و 50 يجيدون الفرنسية و 50 لا يجيدون ايهما . اختير طالب عشوائيا أوجد احتمال :

ا) أن يجيد اللغة الإنجليزية والفرنسية .

ب) أن يجيد الفرنسية علما بأنه لا يجيد الانجليزية .

جـ) أن يجيد لغة واحدة على الأقل .

د) أن يجيد لغة واحدة فقط من الإنجليزية أو الفرنسية .

الفصل الرابع

المتغيرات العشوائية والتوزيعات الاحتمالية
Random Variables and Probability Distribution

● مقدمة ● المتغير العشوائي ● دالة التوزيع الاحتمالي ● التوقع والتباين للمتغير العشوائي ● التوزيعات المنفصلة (المتقطعة) الخاصة ● التوزيعات المتصلة الخاصة ● تمارين

(٤ , ١) مقدمة
(Introduction)

سبق أن درسنا في الفصل الثاني التجربة العشوائية وفراغ العينة والحدث واحتمالات حدوث هذه الاحداث . في هذا الفصل لا يكون اهتمامنا بدراسة نقاط فراغ العينة ولكن يكون اهتمامنا منصبا على قيم عددية مرتبطة بنقاط فراغ العينة . هذه القيم العددية تسمى بالمتغير العشوائي . سندرس في هذا الفصل المتغيرات العشوائية وتوزيعاتها وتوقعاتها وتبايناتها سواء من كان منها متقطعا أو متصلا .

(٤ , ٢) المتغير العشوائي
(Random Variable)

يعرف المتغير العشوائي بصورة مبسطة بأنه دالة ذات قيم عددية حقيقية معرفة على فراغ العينة مجاله فراغ العينة S ومجاله المقابل X(s) هو مجموعة

جزئية من مجموعة الأعداد الحقيقية R وينقسم إلى قسمين هما :

أ) المتغير العشوائي المتقطع (المنفصل) (discrete random variable) .

ب) المتغير العشوائي المتصل (المستمر) (continuous random variable) .

وسنتناول كل منهما بالشرح والأمثلة كما يلي :

(١ , ٢ , ٤) المتغير العشوائي المتقطع (المنفصل)

يقال إن المتغير العشوائي متقطع أو منفصل إذا كان يأخذ قيما تنتمي إلى مجموعة محدودة أو معدودة ، أي يأخذ قيما منفصلة ومتباعدة عن بعضها البعض . من أمثلة المتغير العشوائي المتقطع عدد الصور التي تظهر عند قذف قطعة نقود n من المرات، مقدار الرقم الذي يظهر عند رمي حجر النرد مرة أو أكثر، عدد الوحدات المعيبة عند سحب عينة n من إنتاج إحدى الآلات، عدد حوادث المرور في أحد الشوارع في مدينة ما خلال شهر معين . . . الخ . وسوف نرمز للمتغيرات العشوائية بالحروف الكبيرة .... ,X, Y, Z وقيم هذه المتغيرات بالحروف الصغيرة ... ,x, y, z .

مثال (١ , ٤)

ألقيت قطعة عملة متزنة ثلاث مرات . أوجد القيم الممكنة للمتغيرات العشوائية التالية :

ا) X = عدد مرات ظهور الصورة .

ب) Y = مربع عدد مرات ظهور الصورة .

جـ) Z = الفرق بين عدد مرات ظهور الصورة وعدد ظهور الكتابة .

الحل

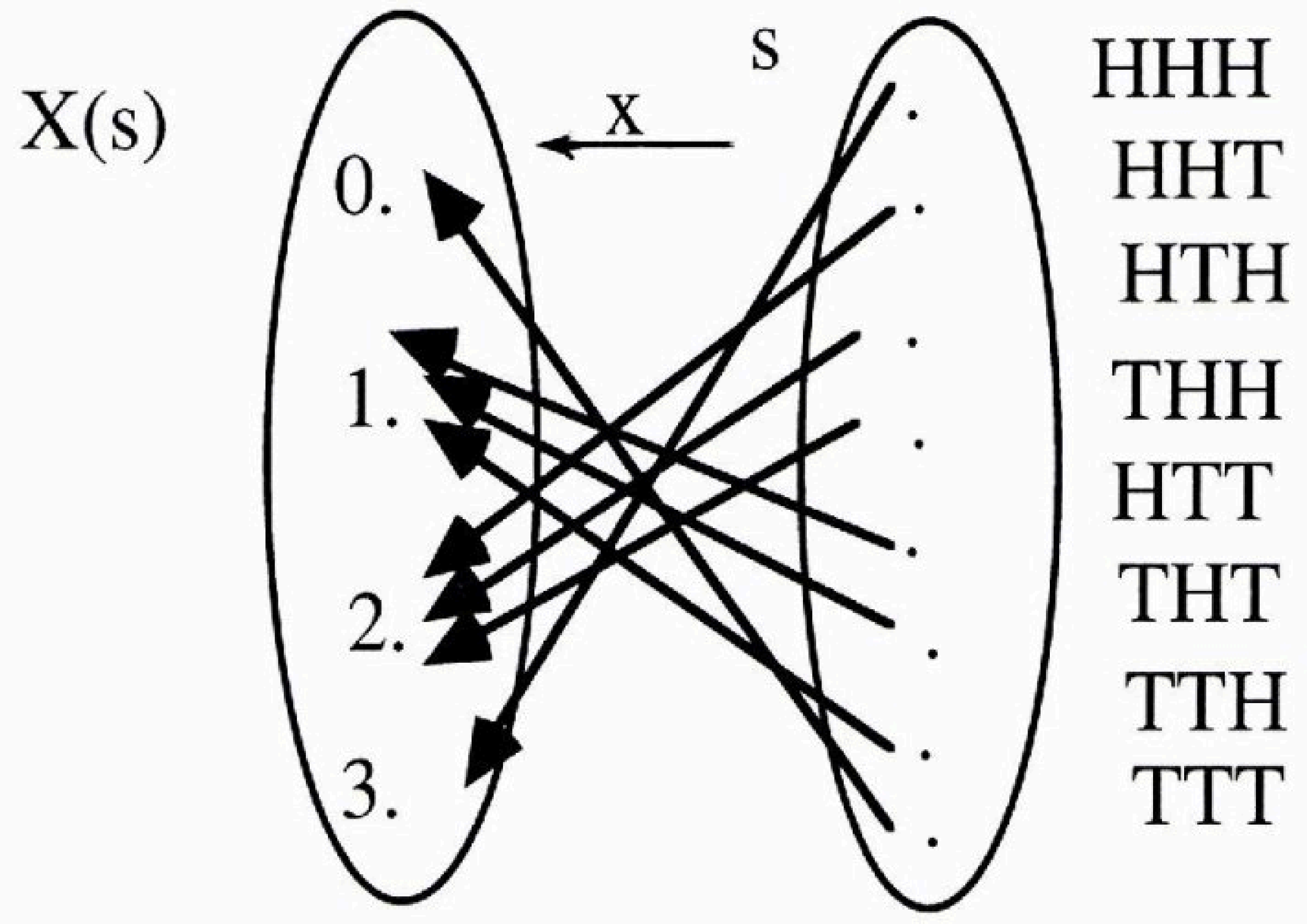
فراغ العينة S هو:

$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

ا) المتغير العشوائي X قيمه الممكنة (مجاله المقابل) X(s) هي:

$$X(s) = \{0,1,2,3\}$$

ويوضح ذلك بالرسم في شكل (١ , ٤) كما يلي :



شكل (١ , ٤)

حيث إن $X(HHH) = 3$

$X(HHT) = 2,\ X(HTH) = 2,\ X(THH) = 2$
$X(TTH) = 1,\ X(THT) = 1,\ X(HTT) = 1$
$X(TTT) = 0$

ب) والقيم الممكنة Y(s) للمتغير العشوائى Y هي :

$$Y(s) = \{0,1,4,9\}$$

جـ) بالمثل القيم الممكنة للمتغير العشوائي Z هي:

$Z(s) = \{ -3, -1, 1, 3\}$

أي أن

$Z(HHH) = 3 - 0 = 3$

$Z(HHT) = 2 - 1 = 1, Z(HTH) = 1, Z(THH) = 1$

$Z(TTH) = 1 - 2 = -1, Z(THT) = -1, Z(HTT) = -1$

$Z(TTT) = 0 - 3 = - 3$

(٤ , ٢ , ٢) المتغير العشوائي المتصل (المستمر)

يقال إن المتغير العشوائي متصلا أو مستمرا إذا كان يأخذ قيما تنتمي إلى مجموعة غير محدودة (لا نهائية) ، أي أنه يأخذ قيم داخل فترة (a,b) من الأعداد الحقيقية R . مثال على ذلك، عمر مصباح في فترة (0,2) من السنوات، أي أن يوجد عدد لا نهائي من الأعمار لهذا المصباح خلال هذه الفترة وكذلك طول أو وزن أحد الطلاب ، الدخل الشهري للأسرة . . . الخ .

لاحظ الاحتمال للمتغير المتصل عند نقطة يساوي صفرا .

(٤ , ٣)دالة التوزيع الاحتمالي
Probability Distribution

عند دراسة المتغير العشوائى X فإنه ليس كافيا معرفة القيم الممكنة للمتغير العشوائى X ولكن قد يكون من المرغوب معرفة السلوك الاحتمالي لهذه المتغيرات العشوائية ، مما يتطلب دراسة دالة التوزيع التراكمي ودالة التوزيع الاحتمالي والتي تنقسم إلى قسمين هما :

أولا: دالة الكتلة الاحتمالية للمتغير العشوائي المنفصل.

ثانيا : دالة الكثافة الاحتمالية للمتغير العشوائى المتصل.

وسنتناول كل منهما بالشرح والأمثلة كما يلي :

(١ , ٣ , ٤) دالة الكتلة الاحتمالية (Probability mass function)

إذا كان لدينا متغير عشوائي متقطـع X يأخذ القيم $x_1, x_2, \ldots$ فإن دالة الكتلة الاحتمالية $f_X(x_i)$ تعرف كما يلي :

$$f_x(x_i) = \begin{cases} P(X = x_i), & X = x_i; \quad i = 1, 2, \cdots \\ 0 & X \neq x_i \end{cases} \tag{٤ , ١}$$

مثال (٤ , ٢)

أوجد قيم دالة الكتلة الاحتمالية $f_X(x)$ للمتغير العشوائي X = عدد ظهور الصورة في مثال (٤ , ١) السابق .

الحل

حيث إن القيم الممكنة للمتغير العشوائي هـي $X(s) = \{0,1,2,3\}$. نحسب $f_X(0)$ ، $f_X(1)$ ، $f_X(2)$ ، $f_X(3)$ للقيم الممكنة كما يلي :

$$f_X(0) = P(X = 0) = P(\{TTT\}) = \frac{1}{8}$$

$$f_X(1) = P(X = 1) = P(\{HTT, THT, TTH\}) = \frac{3}{8}$$

$$f_X(2) = P(X = 2) = P(\{HHT, HTH, THH\}) = \frac{3}{8}$$

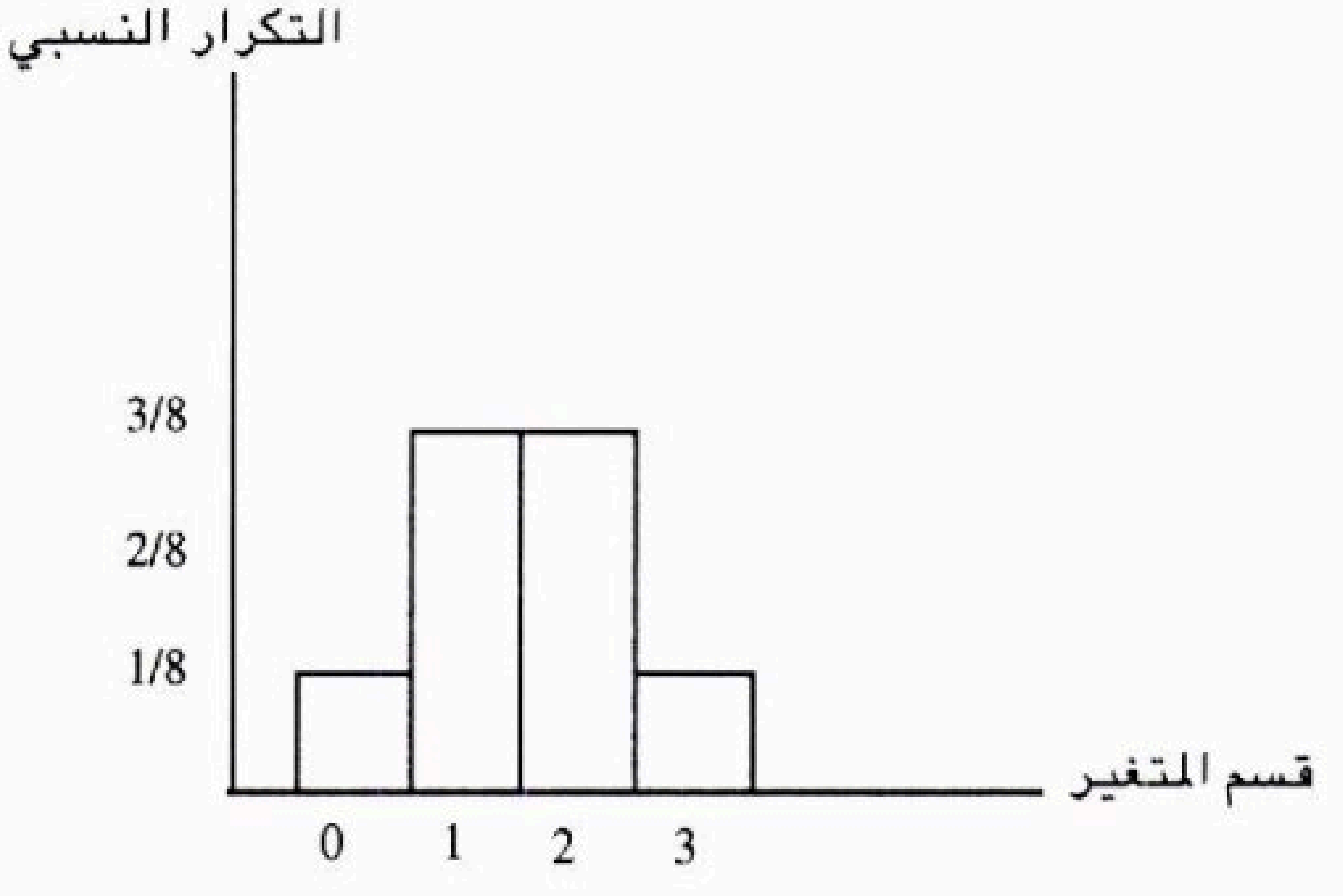
$$f_X(3) = P(X = 3) = P(\{TTT\}) = \frac{1}{8}$$

ويمكن تلخيص النتائج السابقة في جدول (١ , ٤) كما يلي :

قيم المتغير X	0	1	2	3
دالة الكتلة $f_X(x)$	1/8	3/8	3/8	1/8

جدول (١ , ٤)

ويمكن أن تمثل بيانات جدول (١ , ٤) بيانيا بالمدرج التكراري النسبي شكل (٢ , ٤) كما يلي :



شكل (٢ , ٤)

ملاحظة :

يوجد خاصيتان مهمتان لدالة الكتلة الاحتمالية هما :

$$\text{(i)} \quad f_X(x) \geq 0$$

$$\text{(ii)} \quad \sum_{\text{all } X_i} f_X(x_i) = 1$$

مثال (٣ , ٤)

تحقق من أن خاصيتي دالة الكتلة الاحتمالية محققة في مثال (١ , ٤) الخاص بقطعة العملة المتزنة .

الحل

الخاصية (i) محققة لأن

$$f_X(0) = \frac{1}{8} > 0,\ f_X(1) = \frac{3}{8} > 0,\ f_X(2) = \frac{3}{8} > 0,\ f_X(3) = \frac{1}{8} > 0$$

الخاصية (ii) محققة لأن

$$\sum_{\text{all } X_i} f_X(x) = f(0) + f(1) + f(2) + f(3)$$

$$= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$$

(٢ , ٣ , ٤) دالة الكثافة الاحتمالية (Probability density function)

لأي متغير عشوائي متصل توجد دالة كثافة احتمالية بحيث إن المساحات تحت المنحنى لهذه الدالة يمثل الاحتمالات المناظرة للفترات التابعة لها على المحور الأفقي ، أي أن دالة الكثافة الاحتمالية إذا تكاملت من a إلى b أي على الفترة (a,b) فإن الناتج هو احتمال أن المتغير العشوائي X سوف يأخذ أي قيمة في الفترة (a,b) . وتعرف دالة الكثافة $f_X(x)$ كما يلي :

$$P(a \le X \le b) = \int_a^b f_X(x)dx \qquad (٢ , ٤)$$

ولدالة الكثافة الاحتمالية $f_X(x)$ خاصيتان مثل دالة الكتلة الاحتمالية هما :

$$\text{(i)} \quad f_X(x) \geq 0$$

$$\text{(ii)} \quad \int_{-\infty}^{\infty} f_X(x) = 1$$

مثال (٤ , ٤)

كانت دالة الكثافة الاحتمالية $f_X(x)$ للمتغير العشوائي X هي:

$$f_X(x) = \begin{cases} c\,e^{-3x} & \text{for } x > 0 \\ 0 & \text{خلاف ذلك} \end{cases}$$

أوجد الثابت c ثم احسب قيمة الاحتمال للمتغير X في الفترة $[\frac{1}{2}, 1]$

الحل

نستخدم الخاصية (ii) لدالة الكثافة لحساب قيمة الثابت c كما يلي :

$$\int_{-\infty}^{\infty} f_x(x) = P(X \leq x) = \int_0^{\infty} C\ e^{-3x} dx = c\left[\frac{e^{-3x}}{-3}\right]_0^{\infty} = 1 \Rightarrow \frac{c}{3} = 1$$

أي أن $c = 3$ وأن الاحتمال في الفترة (0.5,1) هو

$$P(0.5 \leq X \leq 1) = \int_{0.5}^{1} 3\,e^{-3x}\,dx = -\,e^{-3} + e^{-1.5} = 0.173$$

ملاحظة

تعرف دالة التوزيع التراكمي $F_X(x)$ للمتغير المتصل X كما يلي:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(x)\,dx\ ;\ -\infty < X \leq x \qquad (٤ , ٣)$$

دالة التوزيع التراكمي (Comulative distribution)

تعرف دالة التوزيع التراكمي بأنه إذا كان متغير X يأخذ القيم $x_1, x_2, \dots$ دوال كتلتها المناظرة هي $f(x_1), f(x_2), \dots$ فإن دالة التوزيع التراكمي للمتغير X هي :

$$F_X(x) = P(X \le x)$$

أي أن

$$F_X(x) = \sum_{x_i \le x} f_X(x_i) \quad ; \quad -\infty < X < \infty \qquad (٤ , ٤)$$

يوجد ثلاثة خصائص مهمة لدالة التوزيع التراكمي $F_X(x)$ نلخصها كما يلي :

$$\text{(i)} \quad F_X(x) = P(X \le x)$$

$$F_X(-\infty) = P(X \le -\infty) = 0$$

$$F_X(\infty) = P(X \le +\infty) = 1$$

$$\text{(ii)} \quad F_X(a) \le F_X(b) \quad ; \quad a < b$$

أي أن دالة التوزيع التراكمي متزايدة (غير متناقصة) .

$$\text{(iii)} \quad 0 \le F_X(x) \le 1$$

أي أن حالة التوزيع التراكمي محصورة بين الصفر والواحد الصحيح .

مثال (٤ , ٥)

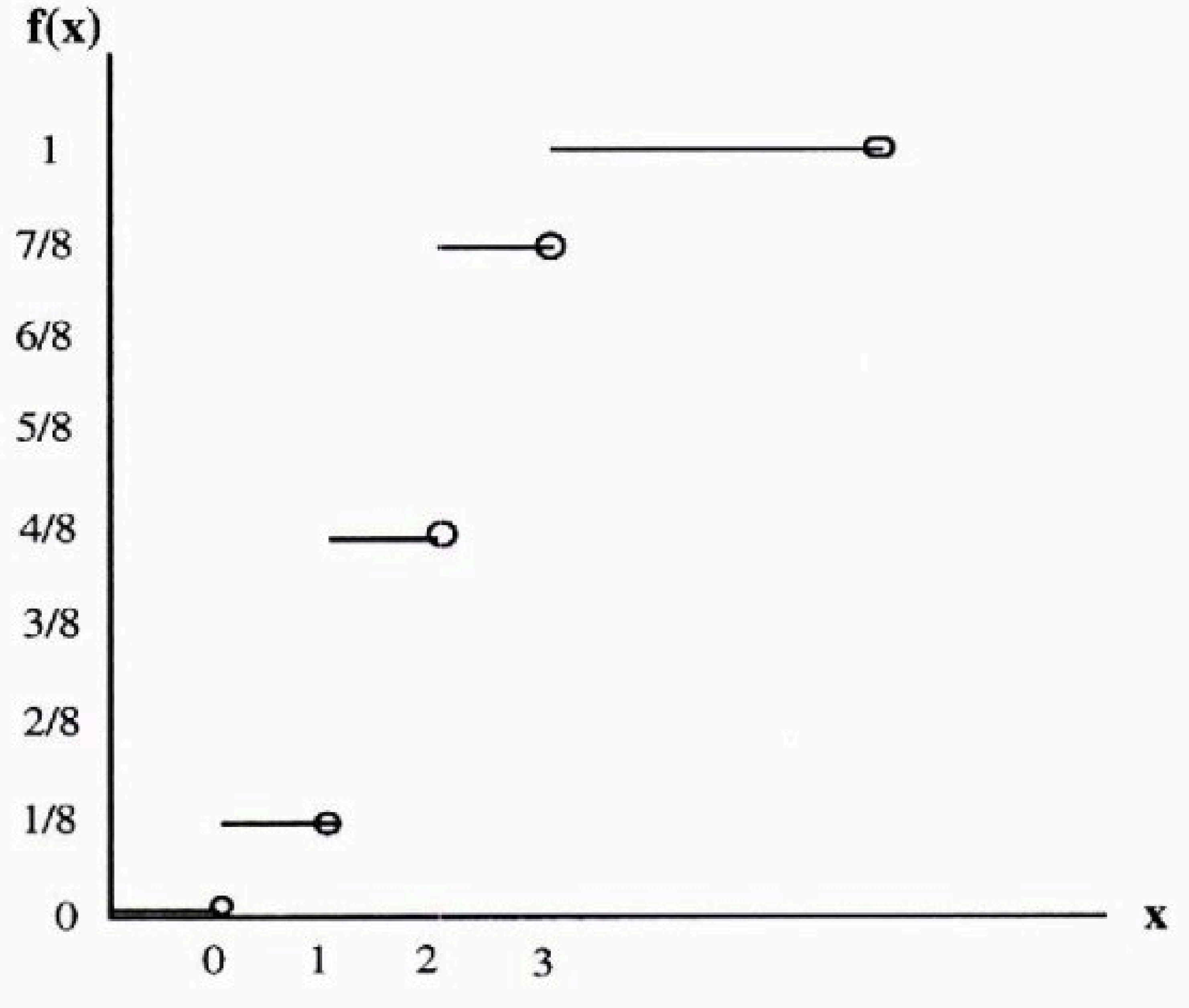
احسب دالة التوزيع التراكمي في مثال (٤ , ٢) السابق الخاص بالعملة المتزنة .

الحل

بتطبيق خواص دالة التوزيع التراكمي $F_X(x)$ يمكن تلخيص الحل كما يلي:

$$F_X(X)=\begin{cases}0 & -\infty \le x<0\\ \frac{1}{8} & 0\le x<1\\ \frac{4}{8} & 1\le x<2\\ \frac{7}{8} & 2\le x<3\\ 1 & 3\le x<\infty\end{cases}$$

ونوضح $F_X(x)$ بيانيا كما في الشكل (٣,٤) كما يلي :



شكل (٣,٤)

(٤ , ٤) التوقع والتباين للمتغير العشوائي
Expectation and Variance of the Random Variable

يوجد عدة خواص تميز التوزيعات الاحتمالية . فمثلا قد لا نهتم بمعرفة كل التفاصيل عن التوزيع الاحتمالي لعمر مصباح كهربائي لماركة معينة ، ونكتفي بمعرفة متوسط وتباين توزيع عمر هذا المصباح مثلا . والتوقع للتوزيع هو مقياس لمتوسط القيم التي يأخذها المتغير العشوائي والتباين ثم الانحراف المعياري هو مقياس درجة تجانس القيم التي يأخذها المتغير العشوائي . وسندرس التوقع والتباين في حالتي المتغير العشوائي المتقطع والمتصل كما يلي :

(١ , ٤ , ٤) التوقع والتباين للمتغير العشوائي المتقطع

إذا كان لدينا متغيرا عشوائيا X يأخذ القيم المنتهية $x_1, x_2, ..., x_n$ ودالة كتلته الاحتمالية المناظرة هي $f(x_1), f(x_2), ..., f(x_n)$ فإن التوقع للمتغير العشوائي X أو المتوسط ويرمز له بالرمز μ_X أو الرمز E(X) يحسب كما يلي:

$$\mu_X = E(X) = x_1 f(x_1) + x_2 f(x_2) + ... + x_n f(x_n)$$

أي أن

$$\mu_X = E(X) = \sum_{i=1}^{n} x_i f(x_i) \qquad (٤ , ٦)$$

والتباين للمتغير العشوائي X ويرمز له بالرمز σ_X^2 أو الرمز V(X) يعطى كما يلي:

$$\sigma_X^2 = V(X) = E(X - \mu_X)^2$$

أو بصيغة أخرى

$$\sigma_X^2 = V(X) = E(X^2) - \mu_X^2$$

أي أن

$$\sigma_X^2 = V(X) = \sum_{i=1}^{n} x_i^2 f(x_i) - \mu_X^2 \qquad (٤ , ٧)$$

مثال (٤ , ٦)

احسب التوقع والتباين ثم الانحراف المعياري σ_X في مثال (٤ , ٢) السابق الخاص بالعملة المتزنة .

الحل

وللسهولة نكون جدولا للحل كما يلي :

x	f(x)	xf(x)	$x^2f(x)$
0	1 / 8	0	0
1	3 / 8	3 / 8	3 / 8
2	3 / 8	6 / 8	12 / 8
3	1 / 8	3 / 8	9 / 8
Σ		1.5	3

فإن التوقع μ_X من الجدول هو:

$$\mu_X = \sum_{x=0}^{3} x\, f(x) = 1.5$$

والتباين σ_X^2 من الجدول أيضا هو :

$$\sigma_X^2 = \sum_{x=0}^{3} x^2 f(x) - \mu_X^2 = 3 - (1.5)^2 = 0.75$$

والانحراف المعياري σ_X هو:

$$\sigma_X = \sqrt{0.75} = 0.81$$

بعض خواص التوقع والتباين :

(i) توقع المقدار $a X \pm b$ حيث a,b ثوابت يعطى كما يلي :

$$E(a X \pm b) = a E(X) \pm b$$

(ii) توقع المقدار الثابت c هو المقدار الثابت نفسه

$$E(c) = c$$

(iii) إذا كان المتغير العشوائي X دالة $G(X)$ ودالة كتلته $f(x)$ فإن

$$E(G(X)) = \sum_{\text{all } x} G(x)\, f(x)$$

(iv) تباين المقدار c يساوي صفرا أي أن :

$$V(c) = 0$$

(v) تباين المقدار $a X \pm b$ حيث إن a,b مقداران ثابتان هو :

$$V(a X \pm b) = a^2\, V(X)$$

(٢ , ٤ , ٤) التوقع والتباين للمتغير العشوائى المتصل

يمكن حساب التوقع والتباين للمتغير المتصل X كما سبق حسابه للمتغير العشوائى المتقطع وذلك باستبدال العلامة Σ التي ترمز للمجموع بالعلامة $\int$ التي ترمز للتكامل وبذلك فإن التوقع والتباين كما يلي:

$$\mu_X = E(X) = \int x\, f(x) \qquad (٤ , ٧)$$

$$\sigma_X^2 = V(X) = \int x^2\, f(x) - \mu_X^2 \qquad (٤ , ٨)$$

مثال (٤ , ٧)

إذا كان لدينا دالة كثافة احتمالية $f_X(x)$ كما يلي:

$$f_x(x)=\begin{cases}\frac{2}{9}x & 0\le x\le 3\\ 0 & \text{خلاف ذلك}\end{cases}$$

أوجد التوقع μ والتباين σ^2 .

الحل

التوقع μ هو:

$$\mu=\int_{-\infty}^{\infty}x\,f(x)=\int_0^3 x\,(\frac{2}{9}x)\,dx=[\frac{2}{27}x^3]_0^3=2$$

والتباين σ^2 هو:

$$\sigma^2=\int_{-\infty}^{\infty}x^2 f(x)-\mu^2=\int_0^3 x^2(\frac{2}{9}x)dx-(2)^2=[\frac{1}{18}x^4]_0^3-4=0.5$$

والانحراف المعياري σ هو

$$\sigma=\sqrt{0.5}=0.7$$

(٥ , ٤) التوزيعات المنفصلة الخاصة

(Discrete distributions)

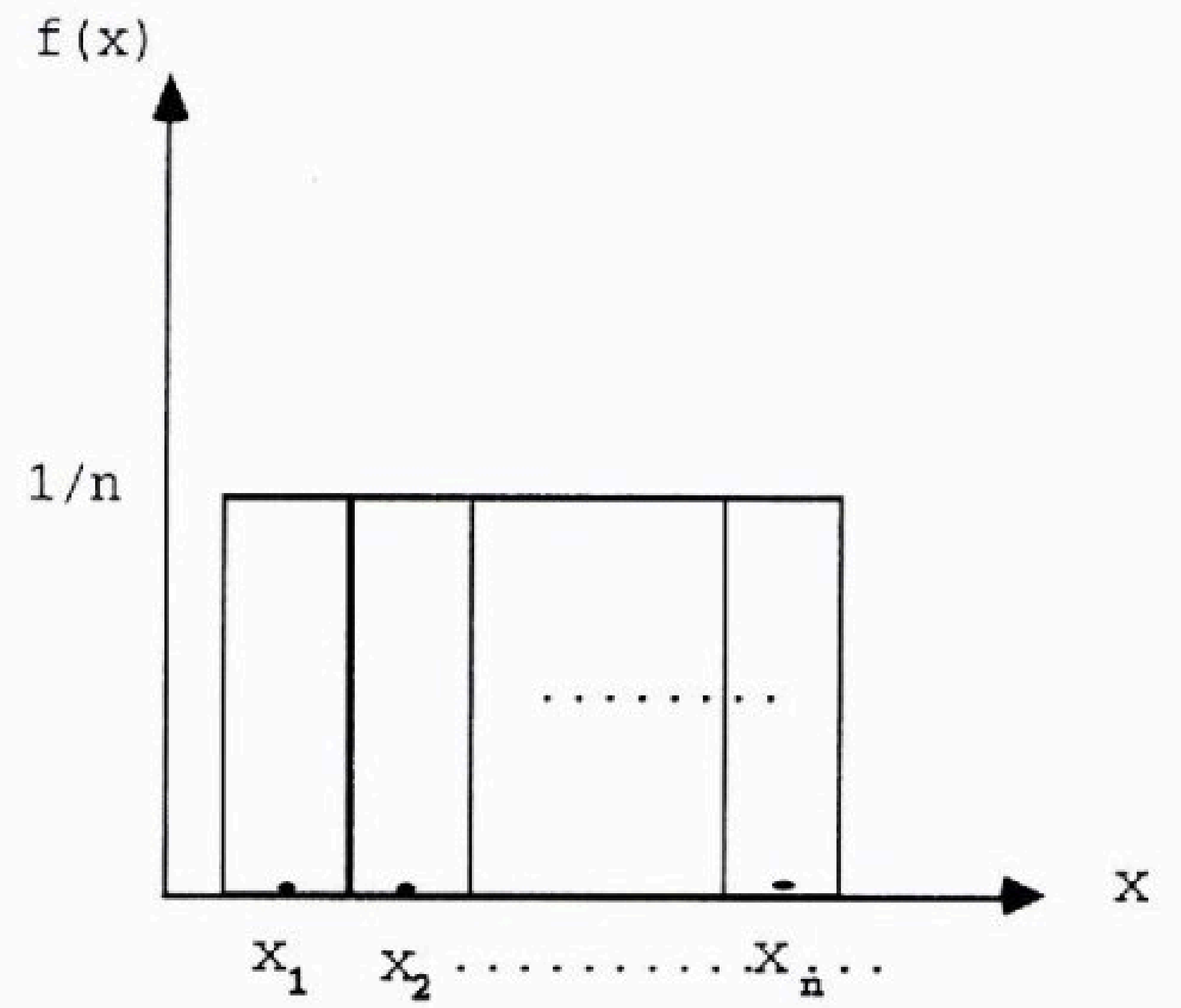
بعد دراسة المتغير العشوائي المنفصل والمتصل بصورة عـامـة. سـوف نستعرض فيما يلي بعض التوزيعات الخاصة مثل التوزيع المنتظم المتقطع وتوزيع برنولي وتوزيع ذي الحدين وتوزيع فوق الهندسي وتوزيع بواسون، وذلك بعرض دالة كتلة كل توزيع وتوقعه وتباينه وبعض الأمثله كما يلي:

(٤ , ٥ , ١) التوزيع المنتظم المتقطع (Discrete uniform distribution)

هو التوزيع التي تأخذ فيه جميع قيم المتغير العشوائي X الاحتمال نفسه . وأن دالة كتلته الاحتمالية $f_X(x)$ كما يلي :

$$f_X(x) = \begin{cases} \dfrac{1}{n}, & x = x_1, x_2, \dots, x_n \\ 0 & \text{خلاف ذلك} \end{cases} \qquad (٤ , ٩)$$

ويمثل المدرج التكراري النسبي للتوزيع المنتظم المتقطع بيانيا باستخدام دالة كتلته (٤ , ٩) بالشكل (٤ , ٤) كما يلي :



شكل (٤ , ٤)

مثال (٤ , ٨)

إذا كان المتغير العشوائي X يمثل الرقم الذي يظهر على الوجه الأعلى عند

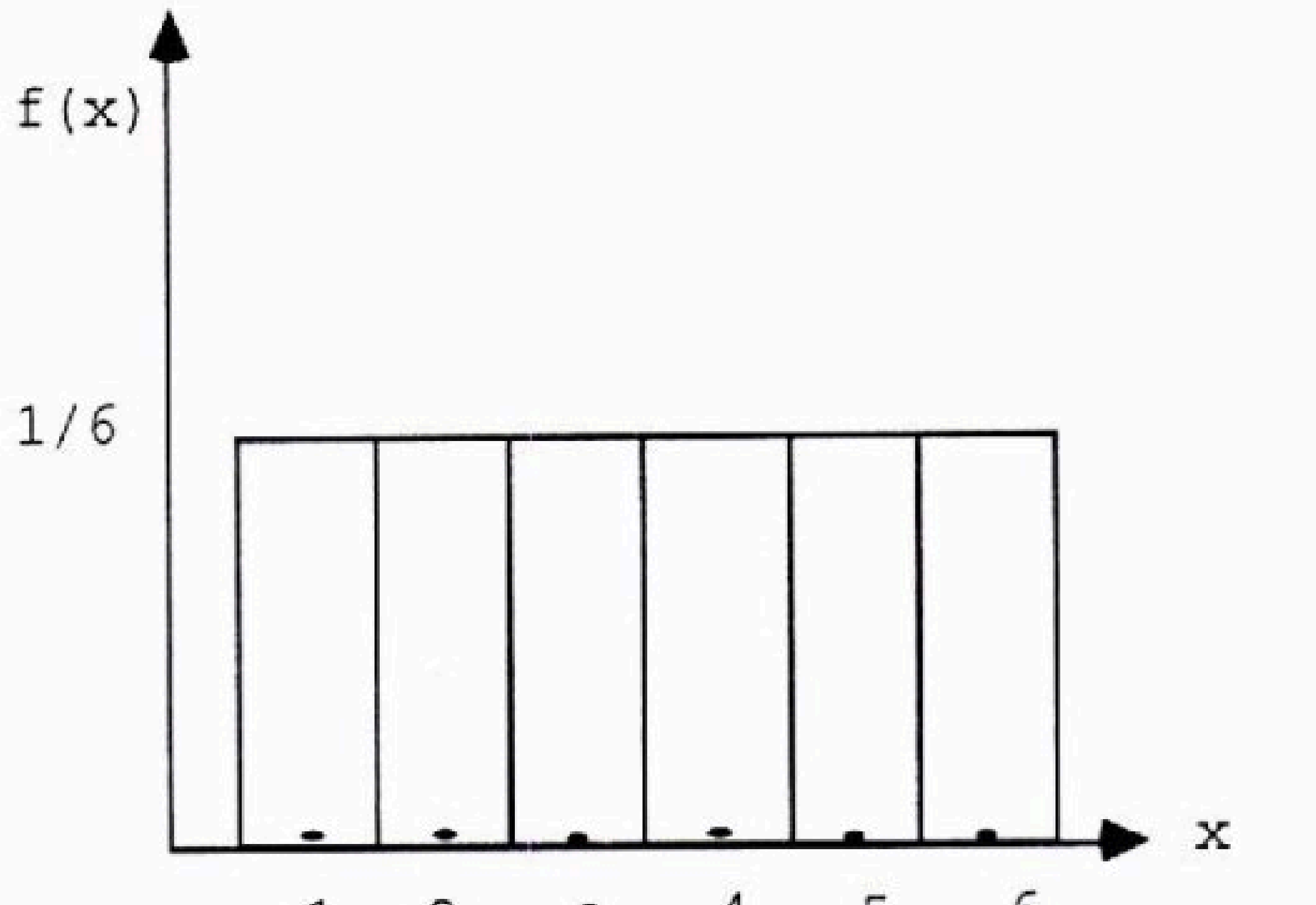
رمي حجر نرد متجانس مرة واحدة . اوجد دالة كتلته الاحتمالية ثم مثلها بيانيا .

الحل

دالة الكتلة الاحتمالية $f_X(x)$ هي :

$$f_X(x) = \begin{cases} \frac{1}{6}, & x = 1,2,3,4,5,6 \\ 0 & \text{خلاف ذلك} \end{cases} \qquad (٤,١٠)$$

ونمثل العلاقة (٤,١٠) السابقة بالمدرج التكراري النسبي شكل (٤,٥) كما يلي :



شكل (٤,٥)

(٤,٥,٢) توزيع برنولي (Bernulli distribution)

يصف توزيع برنولي التجارب العشوائية التي يكون نواتجها ناتجين فقط الناتج الأول يسمى نجاح ويرمز له بالواحد (1) والثاني فشل ويرمز له بالصفر (0)

ويكــون احتمال النجــاح ثابتا ومستقــلا عند تكــرار هذه التجربة العشوائـيــة أكــثر من مرة وتسمى كل تجربة عشوائية بمحاولة برنولي (Bernulli trial) . فمثلا إذا كان المتغير العشوائي (X = عدد مرات ظهور الصورة) عند رمي قطعة عملة مرة واحدة فإن القيم الممكنة للـمـتـغـيـر X هـي $X(s) = \{0,1\}$ باحتمال نجــاح $f_X(1)=P(X=1)=p$ واحتمال فشل $f_X(0)=P(X=0)=q$ وتمثل دالة كتلته الاحتمالية $f_X(x)$ كما يلي :

$$f_X(x) = \begin{cases} \binom{1}{x} p^x q^{1-x}, & x = 0,1 \\ 0 & \text{خلاف ذلك} \end{cases} \qquad (٤,١١)$$

ويتحقق شرطي دالة الكتلة الاحتماليــة وهـمــا $\Sigma f(x) = 1$ و $f(x) \geq 0$ للعلاقــة (٤,١١) كما يلي :

الشرط الأول هو :

$$f_X(0) = \binom{1}{0} p^0 q^{1-0} = q > 0$$

$$f_X(1) = \binom{1}{1} p^1 q^{1-1} = P > 0$$

ويكون الشرط الثاني هو :

$$f_X(0) + f_X(1) = p + q = 1 \qquad (٤,١٢)$$

من العلاقة (٤,١٢) نحسب قيمة q وهي

$$q = 1 - p$$

(٤,٥,٣) توزيع ذي الحدين (الثنائي) (Binomial distribution)

نفرض أن تجربة عشوائية تتكون من تكرار محاولة برنــولــي عــددا n من المرات وفي المحاولة المكررة يظل كل من احتمال النـجــاح p واحتمال الفشل q مقدارا ثابتا في جميع المحاولات . يصف توزيع ذي الحدين جميع المجتمعــات التي يمكن تقسيم مفرداتها إلى صفتين ظهور أحدها نجاحا واحتماله p وعدم ظهور

هذه الصفة مثلا واحتماله q . من هذه المجتمعات سحب عينة n من إنتاج أحد الآلات لدراسة احتمال القطع المعيبة للإنتاج ، سحب عينة من المصابيح بإرجاع من صندوق لدراسة احتمال المصابيح المحروقة ، سحب عينة من الكرات البيضاء بإرجاع لدراسة احتمال الكرات البيضاء في صندوق أو غير البيضاء وهكذا . ومن تكرار محاولة برنولي n من المرات تصبح العلاقة (٤ , ١٢) كما يلي :

$$(p+q)^{n} = 1 \qquad (٤ , ١٣)$$

باستخدام مفكوك ذي الحدين للعلاقة (٤ , ١٣) نحصل على المفكوك التالي :

$$\binom{n}{0}P^{0}q^{n-0} + \binom{n}{1}p^{1}q^{n-1} + \ldots + \binom{n}{x}p^{x}q^{n-x} + \ldots + \binom{n}{n}p^{n}q^{0} = 1$$

(٤ , ١٤) . . .

والعلاقة (٤ , ١٤) تحقق شرطي دالـة الكتلـة الاحتماليـة وهما $\Sigma f(x) = 1$ و $f(x) \quad 0$ وبذلك تصبح حدود المفكوك (٤ , ١٤) السابق تمثل دوال كتلـة $f(x)$ حيث

$$f(0) = \binom{n}{0}p^{0}q^{n-0} > 0,\ f(1) = \binom{n}{1}p\,q^{n-1} > 0, \ldots, f(x) = \binom{n}{x}p^{x}q^{n-x} > 0$$

$$,\cdots, f(n) = \binom{n}{n}p^{n}q^{0} > 0$$

أي أن الشرط الأول لدالة الكتلة محقق وأن

$$f(0) + f(1) + \ldots + f(x) + \ldots + f(n) = 1$$

أي أن الشرط الثاني لدالة الكتلة محققا أيضا .

ومن ذلك يمكن تعريف دالة الكتلة الاحتماليـة $f_X(x)$ لتوزيع ذي الحدين (الثنائي) كما يلي :

$$f_X(x) = \begin{cases} \binom{n}{x} p^{x} q^{n-x}, & x = 0,1,2,\ldots,n \\ 0 & \text{خلاف ذلك} \end{cases} \qquad (٤ , ١٥)$$

ويحسب التوقع والتباين لتوزيع ذي الحدين باستخدام العلاقة (٤ , ١٥) كما يلي :

$$\mu = E(X) = \sum_{x=0}^{n} x \,.\, f_X(x)\,, \quad x = 0,1,\ldots,n$$

$$= \sum_{x=0}^{n} x \,.\, \binom{n}{x} p^{x} q^{n-x}$$

بقليل من الاختصار نحصل على :

$$\mu = E(X) = n\,p \qquad (٤ , ١٦)$$

ويحسب التباين σ_X^2 كما يلي :

$$\sigma_X^2 = V(X) = \sum_{x=0}^{n} x^2 \,.\, \binom{n}{x} p^{x} q^{n-x} - (np)^2$$

وبقليل من الاختصار نحصل على :

$$\sigma_X^2 = V(X) = n\,p\,q \qquad (٤ , ١٧)$$

مثال (٤ , ٩)

إذا ألقيت قطعة عملة متزنة عشر مرات اوجد ما يلي:

ا) اكتب دالة الكتلة الاحتمالية $f(x)$ للمتغير العشوائي X=عدد الصور

ب) احتمال عدم ظهور الصورة .

جـ) احتمال ظهور صورة واحدة على الأقل .

د) احسب التوقع والتباين μ , σ^2 .

الحل

في حالة قطعة العملة المتزنة نجد أن $p = q = \dfrac{1}{2}$

في مثالنا $n = 10$ فإن دالة الكتلة الاحتمالية $f(x)$ هي :

أ) $$f(x) = \binom{10}{x} \left(\frac{1}{2}\right)^{x} \left(\frac{1}{2}\right)^{10-x} , \quad x = 0,1,2,...,10$$

ب) $$f(0) = \binom{10}{0} \left(\frac{1}{2}\right)^{0} \left(\frac{1}{2}\right)^{10-0} = \left(\frac{1}{2}\right)^{10} = \frac{1}{1024}$$

ج) $$P(X \quad 1) = f(1) + f(2) +...) f(10)$$

أو بعبارة أخرى فإن

$$P(X \quad 1) = 1 - f(0)$$
$$= 1 - \frac{1}{1024} = \frac{1023}{1024}$$

د) $$\mu = n\,p = 10\left(\frac{1}{2}\right) = 5$$

$$\sigma^{2} = n\,p\,q = 10\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 2.5$$

(٤ , ٥ , ٤) التوزيع فوق الهندسي (Hypergeometric distribution)

عندما يكون السحب من الصندوق لعدد من المصابيح (محروق، سليم) وكان السحب بدون إرجاع فإننا نلاحظ احتمال نتيجة السحب الثانية تعتمد على نتيجة السحبة الأولى وهكذا لباقي السحبات بدون إرجاع، وبذلك الاحتمال الثابت لكل سحبة غير محقق (محاولة برنولي شرطيها غير محقق). وفي هذه الحالة نجد أن تمثيل توزيع ذي الحدين لاحتمال عدد n من الوحدات يكون غير ممثل للاحتمال الحقيقي ولذلك دعت الحاجة لايجاد توزيع آخر هو توزيع فوق الهندسي الذي لا يشترط ثبات احتمال كل سحبة (محاولة). ونستنتج هذا التوزيع كما يلي.

نفترض أن مجتمع يتكون من صفتين الأولى وعددها a مفردة والثانية عددها b مفردة من مجتمع محدود حجمه N. وسحبت منه عينة n أقل من كل

من a, b وكانت X = عدد المصابيح المحروقة(على سبيل المثال) في صندوق به N مصباح وكان السحب بدون إرجاع . فإن عدد طرق سحب x مصباح محروق و n - x مصباح سليم يساوي$\binom{a}{x}\binom{b}{n-x}$ وأن عدد طرق سحب n مصباح من N مصباح = $\binom{N}{n}$ فإن احتمال الحصول على x مصباح محروق ونرمز له بالرمز $f_X(x)$ وهي تمثل دالة الكتلة الاحتمالية لتوزيع فوق الهندسي كما يلي :

$$f_X(x) = P(X = x) = \begin{cases} \dfrac{\binom{a}{x}\binom{b}{n-x}}{\binom{N}{n}} & x \le a,\ n - x \le b \text{ حيث}, \quad x = 0,1,\ldots,n \\ 0 & \text{خلاف ذلك} \end{cases}$$

(٤ , ١٨) . . .

وإذا كانت النسبة p للمصابيح المحروقة معروفة فإن a = N p وأن المصابيح السليمة هو b = N q وتصبح العلاقة (٤ , ١٨) كما يلي :

$$f_X(x) = P(X = x) = \begin{cases} \dfrac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}} & x \le Np,\ n - x \le Nq \quad \text{حيث} \ , \ x = 0,1,\ldots,n \\ 0 & \text{خلاف ذلك} \end{cases}$$

(٤ , ١٩) . . .

يحسب التوقع μ والتبايـن σ^2 لتوزيع فوق الهندسي باستخدام دالة كتلتـه f(x) المعطاة بالعلاقة (٤ , ١٩) كما يلي :

$$\mu = \sum_{x=0}^{n} x \,.\, f(x) \ , \qquad x = 0,1,2,..,n$$

أي أن

$$\mu = \sum_{x=0}^{n} x \, . \, \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}} \, , \qquad x = 0,1,2,\ldots,n$$

وبقليل من الاختصار نحصل على

$$\mu = n\,p$$

ويحسب التباين σ^2 أيضا كما يلي :

$$\sigma^2 = \sum_{x=0}^{n} x^2 f(x) - \mu^2$$

أي أن

$$\sigma^2 = \sum_{x=0}^{n} x^2 \, . \, \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}} - (n\,p)^2$$

وبعد الاختصار نحصل على تباين توزيع فوق الهندسي كما يلي :

$$\sigma^2 = n\,p\,q\left(\frac{N-n}{N-1}\right) \qquad (٤,٢٠)$$

ملاحظة

إذا كان حجم المجتمع (N) كبيرا جدا فإن التوزيع فوق الهندسي يؤول إلى التوزيع ذي الحدين ويصبح التباين σ^2 المعطى بالعلاقة (٤,٢٠) هو تباين توزيع ذي الحدين المعطى بالعلاقة (٤,١٧) حيث إن

$$\sigma^2 = n\,p\,q\left(\frac{N-n}{N-1}\right) = n\,p\,q\left(1 - \frac{n-1}{N-1}\right) \qquad (٤,٢١)$$

وحيث إن N كبيرا جدا بالنسبة لحجم العينة n أي أن $n < N$ فإن النسبة $\frac{n-1}{N-1}$ يمكن اهمالها وعادة تهمل عندما تكون النسبة $\frac{n}{N}$ أقل من 5% وبذلك تصبح العلاقة (٤,٢١) كما يلي

$$\sigma^2 = n\,p\,q\,(1 - \frac{n-1}{N-1})$$
$$= n\,p\,q\,(1-0) = n\,p\,q$$

وهو تباين توزيع ذي الحدين .

مثال (٤,١٠)

حوض زجاجي لأسماك الزينة به 48 سمكة من بينها 8 سمكات غير ملونة. اختيرت عينة من 5 سمكات أوجد :

ا) دالة الكتلة الاحتمالية لعدد الأسماك غير الملونة .

ب) احتمال العينة المختارة تكون من الأسماك الملونة .

جـ) احتمال العينة المختارة تحتوي على سمكة واحدة غير ملونة .

د) احتمال العينة المختارة تحتوي على سمكة واحدة على الأكثر غير ملونة .

هـ) احسب التوقع والتباين والانحراف المعياري للأسماك غير الملونة.

الحل

نفترض أن المتغير العشوائي X = عدد الأسماك غير الملونة في العينة . من المثال نجد أن $b = 40$ ، $a = 8$ و $n = 5$

$$p = \frac{8}{48} = \frac{1}{6}\ ,\ q = 1 - \frac{1}{6} = \frac{5}{6}$$

فإن دالة الكتلة الاحتمالية $f(x)$ للمتغير X تحسب من العلاقة (٤,١٨) كما يلي :

ا)

$$f_X(x) = \begin{cases} \dfrac{\binom{8}{x}\binom{40}{5-x}}{\binom{48}{5}} & ; \quad x = 0,1,2,..,5 \\ 0 & \text{خلاف ذلك} \end{cases}$$

ب) احتمال العينة كلها أسماك ملونة هو $P(X = 0)$ أي أن

$$f(0) = \binom{8}{0}\binom{40}{5-0} / \binom{48}{5} = 0.38$$

ج) احتمال أن تكون سمكة واحدة فقط غير ملونة في العينة $P(X = 1)$ أي أن

$$f(1) = \binom{8}{1}\binom{40}{5-1} / \binom{48}{5} = 0.427$$

د) احتمال أن تكون سمكة على الأكثر غير ملونة $P(X \quad 1)$ أي أن

$$P(X \quad 1) = f(0) + f(1)$$

$$= 0.38 + 0.427 = 0.807$$

هـ) التوقع والتباين للأسماك غير الملونة هو :

$$\mu = n\,p = 5\,.(\frac{1}{6}) = 0.833 \quad \text{سمكة غير ملونة}$$

$$\sigma^2 = n\,p\,q\,(\frac{N - n}{N - 1}) = 5(\frac{1}{6})(\frac{5}{6})(\frac{48-5}{48-1}) = 0.64$$

ويكون الانحراف المعياري هو :

$$\sigma = \sqrt{0.64} = 0.8 \quad \text{سمكة غير ملونة}$$

(٤,٥,٥) توزيع بواسون (Poisson distribution)

يهتم توزيع بواسون بالأحداث نادرة الحدوث أي عند احتمال حدوث حدث A مثلا $P(A) = p$ صغيرا جدا في مجتمع حجمه N كبيرا جدا . وقد اكتشفه العالم الالماني بواسون (١٧٨٧م) في دراسة المتغير العشوائي X الذي يمثل عدد الجنود الالمان الذين يموتون من رفسة حصان خلال فترة زمنية محدودة. وأصبح لتوزيع بواسون الآن تطبيقات كثيرة في الحياة، نذكر منها حساب الاحتمالات لعدد المكالمات الهاتفية التي تصل إلى سنترال معين خلال دقيقة مثلا واحتمال عدد الأخطاء الموجودة في صفحة قاموس واحتمال عدد حوادث المرور الخطرة في أحد شوارع مدينة ما خلال أسبوع مثلا واحتمال عدد العيوب في طول قماش في إنتاج ما . . . الخ. التجربة العشوائية للمتغير العشوائي X التي تكون دالة كتلته f(x) تمثل توزيع بواسون يجب أن تحقق شروط معينة، فعلى سبيل المثال عدد المكالمات الهاتفية التي تصل إلى سنترال خلال فترة زمنية صغيرة $(t, t + \Delta t)$ يجب أن تحقق الشروط التالية :

ا) احتمال وصول مكالمة هاتفية في فترة زمنية قصيرة $(t, t + \Delta t)$ يساوي $\lambda \Delta t + 0(\Delta t)$ ، حيث λ مقدار ثابت يعتمد على طبيعة الظاهرة.

ب) احتمال وصول أكثر من مكالمة في الفترة الزمنية $(t, t + \Delta t)$ صغيرا جدا ويمكن اعتباره صفرا أي أن $0(\Delta t) = 0$.

جـ) عدد المكالمات التي تصل إلى السنترال في الفترات الزمنية المنفصلة تكون مستقلة.

وبتقسيم فترة زمنية طولها t إلى فترات كل منها Δt وعدد n فترة فإن

$$\Delta t = \frac{t}{n}$$

وأن الفترات مستقلة (محاولة برنولي) واحتمال كل منها مقدار ثابت $\lambda \Delta t = \frac{\lambda t}{n}$

وبتطبيق توزيع ذي الحدين فيكون الاحتمال $P(X = x)$ هو

$$P(X=x)=\binom{n}{x}\left(\frac{\lambda t}{n}\right)^{x}\left(1-\frac{\lambda t}{n}\right)^{n-x}, \qquad x=0,1,\ldots,n$$

وعندما n كبيرة جدا فإن هذا الاحتمال يؤول إلى دالة الكتلة الاحتمالية لتوزيع بواسون $f_X(x)$ كما يلي :

$$f_X(x)=\lim_{n\to\infty}\binom{n}{x}\left(\frac{\lambda t}{n}\right)^{x}\left(1-\frac{\lambda t}{n}\right)^{n-x}, \quad x=0,1,2,\ldots$$

وباعتبار $t = 1$ (وحدة الزمن) فإننا نحصل على :

$$f_X(x)=\begin{cases}\dfrac{\lambda^x}{X!}e^{-\lambda}, & x=0,1,2,\cdots \\ 0 & \text{خلاف ذلك}\end{cases} \qquad (٤,٢٢)$$

ونحسب التوقع والتباين لتوزيع بواسون من العلاقة (٤,٢٢) كما يلي:

$$\mu=E(X)=\sum_{x=0}^{\infty}x\,.\,\frac{\lambda^{x}}{x\,!}\;e^{-\lambda x}=\lambda$$

وتباين توزيع بواسون σ^2 هو :

$$\sigma^2=\sum_{x=0}^{\infty}x^2\,.\,\frac{\lambda^{x}}{x\,!}\;e^{-\lambda}-(\lambda)^2=\lambda$$

والانحراف المعياري σ هو $\sqrt{\lambda}$.

مثال (١١ , ٤)

معدل عدد حوادث السيارات في أحد شوارع مدينة ما هو 2 في الاسبوع .

أوجد

ا) اكتب دالة التوزيع الاحتمالي للمتغير العشوائى X = عدد الحوادث

ب) ما هو احتمال عدم حدوث أي حادث في هذا الشارع خلال أسبوع ما

جـ) ما هو احتمال حدوث حادث واحد فقط .

د) ما هو احتمال حدوث حادث واحد على الأكثر .

هـ) أوجد التوقع والتباين والانحراف المعياري لعدد الحوادث .

الحل

إذا كان المتغير العشوائي X = عدد الحوادث خلال أسبوع في أحد الشوارع فإن

ا) دالة الكتلة $f(x)$ هي (حيث $\lambda = 2$)

$$f(x) = \frac{(2)^{x}}{x!}\, e^{-2}, \quad x = 0,1,2,....$$

ب) احتمال عدم حدوث أي حادث خلال أسبوع $P(X=0)$ أي أن

$$f(0) = \frac{2^{0}}{0!}\, e^{-2} = e^{-2} = 0.1353$$

جـ) احتمال حدوث حادثة واحدة فقط $P(X = 1)$ أي أن

$$f(1) = \frac{2^{1}}{1\,!}\, e^{-2} = 2\, e^{-2} = 0.2706$$

د) احتمال حدوث حادث واحد على الأكثر $P(X\ \ 1)$ أي أن :

$$P(X \le 1) = f(0) + f(1) = e^{-2} + 2\, e^{-2} = 0.4059$$

هـ) التوقع μ والتباين σ^2 لتوزيع بواسون هما :

$$\mu = \sigma^{2} = \lambda = 2 \qquad \text{حادثة}$$

والانحراف المعياري σ هو :

$$\sigma = \sqrt{2} \quad \text{حادثة}$$

مثال (٤,١٢)

في مثال (٤,١٠) السابق احسب الاحتمالات التالية خلال أسبوعين :

أ) دالة الكتلة الاحتمالية إذا كان المتغير العشوائي X = عدد الحوادث خلال اسبوعين.

ب) ما هو احتمال حدوث حادث واحد فقط .

جـ) احسب التوقع والتباين والانحراف المعياري للمتغير X .

الحل

المتغير العشوائي X = عدد الحوادث خلال أسبوعين . إذن يجب أن نجد معدل عدد الحوادث في أسبوعين أي أن :

$$\lambda^* = 2\lambda = 2(2) = 4$$

أ) فإن دالة الكتلة الاحتمالية $f(x)$ هي :

$$f(x) = \frac{4^x}{x!} e^{-4}, \quad x = 0,1,2,....$$

ب) احتمال حدوث حادثة واحدت فقط $P(X = 1)$ هو

$$f(1) = \frac{4}{1!} e^{-4} = 4e^{-4}$$

ج) التوقع μ والتباين σ^2 والانحراف المعياري σ هو :

$$\mu = \sigma^2 = 4$$

$$\sigma = \sqrt{4} = 2\text{حادثة}$$

ملاحظة

يقرب توزيع ذي الحدين إلى توزيع بواسون عندما تكون p صغيرة جدا و n كبيرة جدا وذلك بوضع $n\,p = \lambda$ ويوضح ذلك بالمثال التالي.

مثال (٤ , ١٣)

نسبة المصابين بمرض معين في بلد ما هي 0.001 ما احتمال عدم وجود أي مصاب بهذا المرض في حي يقطنه 5000 نسمة .

الحل

المتغير العشوائي X = عدد المصابين بهذا المرض في الحي

باعتبار X له توزيع ذي الحدين حيث $P=0.001$ ، $n = 5000$ فإن

$$P(X=0) = \binom{5000}{0} (0.001)^{0} (0.999)^{5000} = 0.00672$$

وإذا أردنا استخدام التقريب بتوزيع بواسون حيث

$$\lambda = np = 5000(0.001) = 5$$

$$P(X=0) = \frac{e^{-5}\, 5^{0}}{0!} = 0.00672$$

وتظهر المقارنة بين النتيجتين مدى دقة التقريب

(٤ , ٦) التوزيعات المتصلة الخاصة

(Continiuous Distribution)

سبق أن استعرضنا المتغير العشوائي المنفصل والمتصل ودراسة دالة الكتلة الاحتمالية للمتغير المنفصل ودالة الكثافة الاحتمالية للمتغير المتصل وذلك بصورة

عامة . كذلك درسنا بعض التوزيعات الخاصة للمتغير المنفصل مثل التوزيع المنتظم المتقطع وتوزيع برنولي وتوزيع ذي الحدين وتوزيع فوق الهندسي وتوزيع بواسون . . . . الخ . وفي هذا البند سوف ندرس بعض التوزيعات المتصلة مثل التوزيع الطبيعي وكذلك تقريب توزيع ذي الحدين إلى التوزيع الطبيعي القياسي وتوزيع t وتوزيع مربع كاي وتوزيع F كما يلي :

(١ , ٦ , ٤) التوزيع الطبيعي (المعتدل) Normal distribution

يعتبر التوزيع الطبيعي من أهم التوزيعات الاحتمالية في علم الإحصاء لأنه يمثل كثيرا من الظواهر الطبيعية وغير الطبيعية التي تقابلنا في الحياة العملية مثل الأطوال والأوزان والأعمار ودرجات الحرارة والدخول الشهرية ودرجات الامتحان ، ضغط الدم وأخطاء القياسات ، وغيرها من الظواهر المتصلة .

فإذا كان المتغير العشوائي X مثلا . وكان لمجموعة مشاهداته $x_1, x_2, ..., x_n$ منحنيا تكراريا متماثلا حول محور التناظر المار بمتوسط هذه البيانات . تمكن العالم الألماني جاوس من إيجاد دالة كثافته الاحتمالية f(x) سميت بدالة الكثافة للتوزيع الطبيعي وهي :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (٤ , ٢٣)$$

حيث إن

$$-\infty < x < \infty$$

$$-\infty < \mu < \infty$$

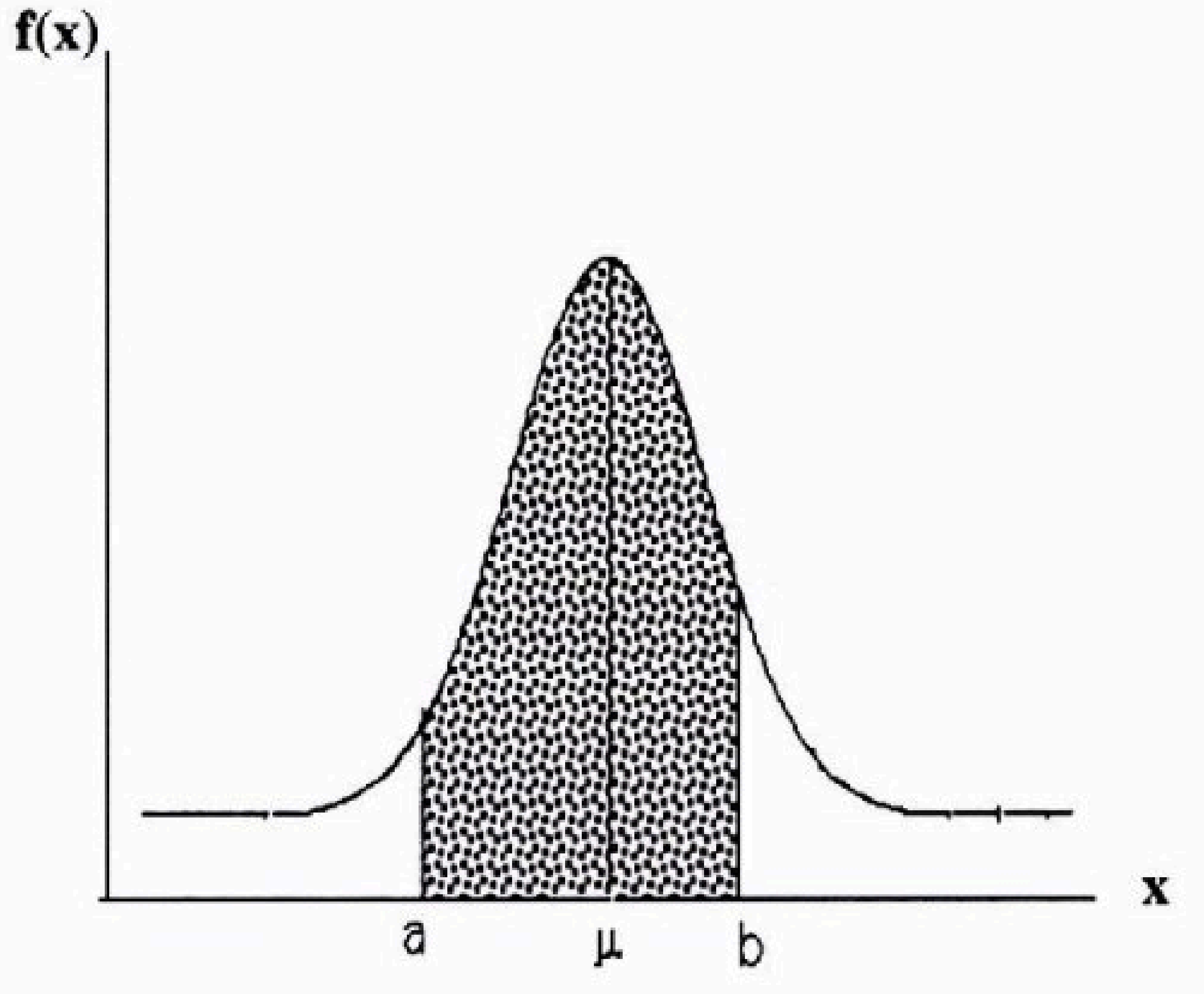
$$0 < \sigma \leq \infty$$

$$\pi = 3.14159 \ldots ; \; e = 2.71828 \ldots$$

واحتمال المتغير العشوائي X أن يقع في الفترة [a,b] هو

$$P(a \leq X \leq b) = \int_a^b f(x)dx \qquad (٤ , ٢٤)$$

ويوضح الاحتمال المعطى بالعلاقة (٤,٢٤) بالمساحة المظللة شكل (٤,٦) كما يلي



شكل (٤,٦)

نلخص في النقاط التالية بعض من خصائص منحنى التوزيع الطبيعي:

ا) له قيمة واحدة ويشبه الناقوس ومتماثل حول المتوسط μ .

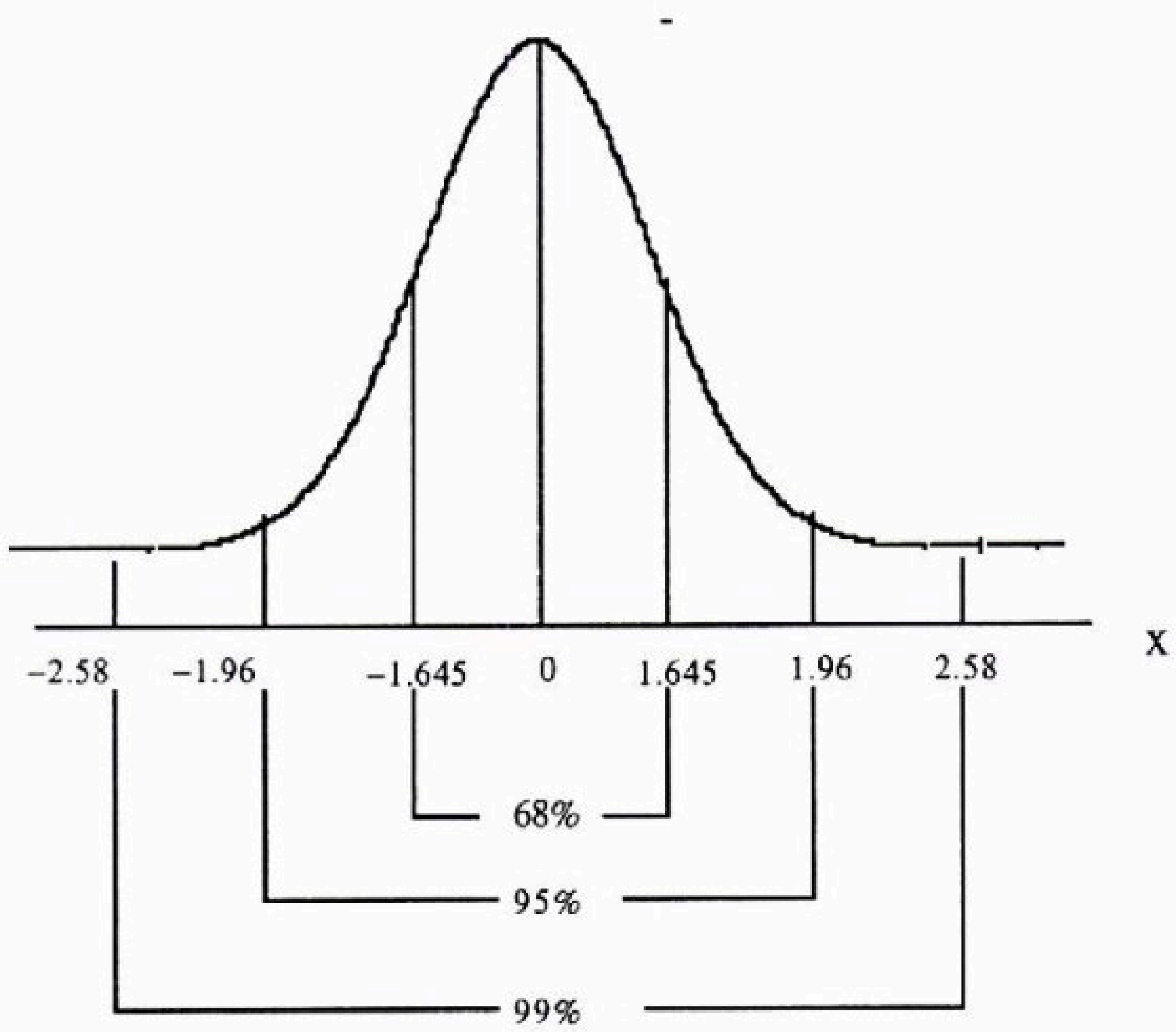
ب) له نقطتا انقلاب عند القيمتين $x = \mu \pm \sigma$.

ج) 68% من قيم المتغير X داخل الفترة $(\mu - \sigma , \mu + \sigma)$.

د) 95% من قيم المتغير X داخل الفترة $(\mu - 1.96\sigma, \mu + 1.96\sigma)$.

هـ) 99% من قيم المتغير X داخل الفترة $(\mu - 2.58\sigma, \mu + 2.58\sigma)$

و) يتقارب المنحنى من الصفر على الجهتين من $x = \mu$ عندما $x \to -\infty$ وعندما $x \to +\infty$. ويوضح ذلك بالرسم شكل (٤,٧) كما يلي :



شكل (٤ , ٧)

(٤ , ٦ , ٢) التوزيع الطبيعي القياسي (Standard normal distribution)

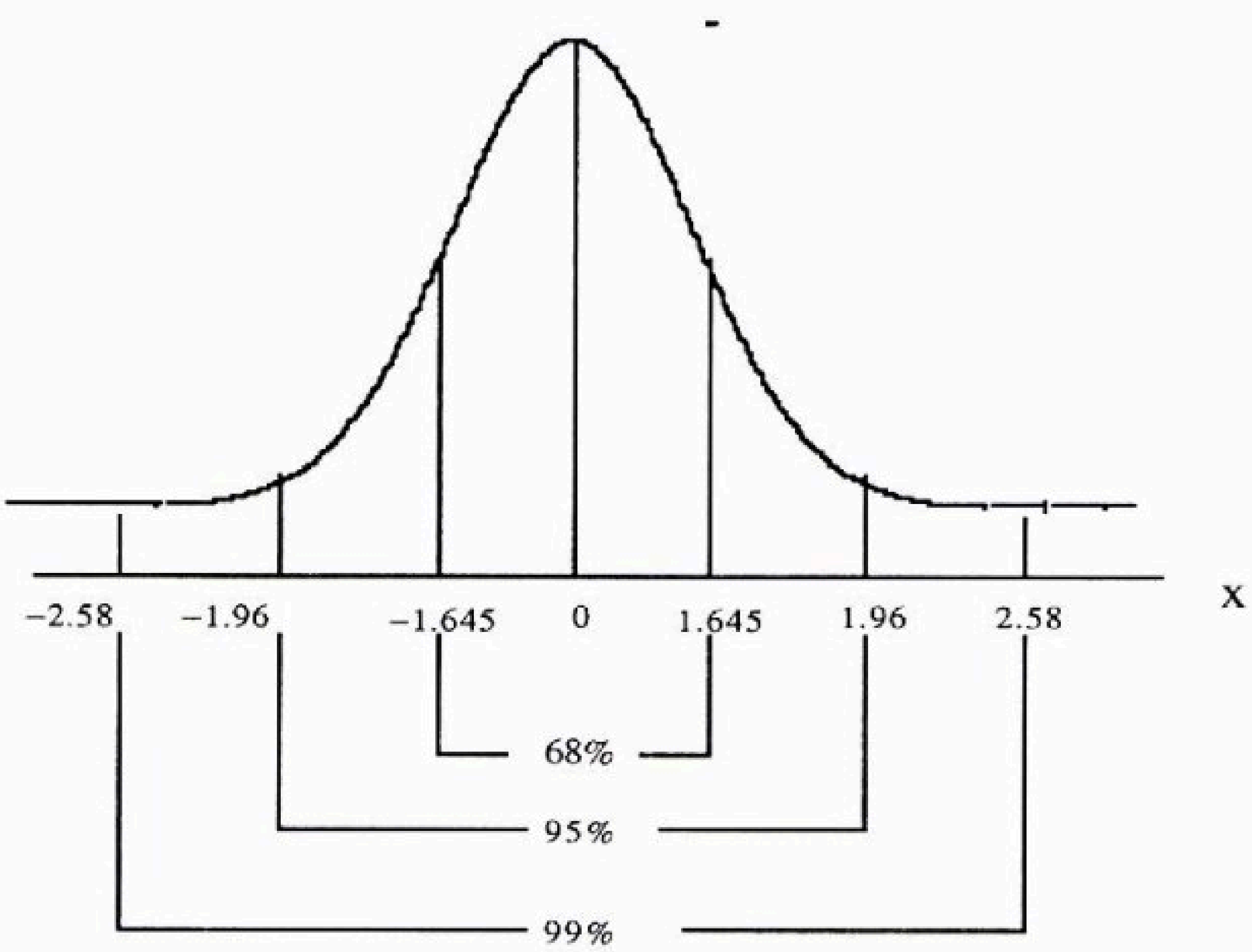
بعد دراسة التوزيع الطبيعي للمتغير العشوائـي X الذي متوسطه μ وتباينه σ^2 ويكتب $N(\mu, \sigma^2)$ حيث المعلمتان μ, σ^2 تؤثران على مكان التوزيع وعلى مقدار تشتته . فقد وجد توزيع نظري متوسطه صفرا وتباينه 1 ويكتب $N(0,1)$ يمكن تحويل جميع التوزيعات الطبيعية إليه . حيث تنطبق عليه نفس الاحتمال في شكل (٤ , ٧) السابق بعد وضع $\mu = 0, \sigma = 1$ وذلك بتحويل المتغير X إلى المتغير Z حيث $Z = \dfrac{X - \mu}{\sigma}$ فإن دالة الكثافة الاحتمالية للمتغير Z هي $\phi(z)$ تعطى بالعلاقة (٤ , ٢٥) كما يلي :

$$\phi(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}} \quad ; \quad -\infty < z < \infty \qquad (٤ , ٢٥)$$

وأن دالة احتماله التراكمية $\Phi(z)$ تمثل بالمساحة تحت المنحنى الطبيعي القياسي من $-\infty$ إلى النقطة Z ويعبر عن ذلك بالعلاقة (٤ , ٢٦) كما يلي:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}}\, dt \qquad (٤ , ٢٦)$$

ويوضح الاحتمال للمتغير Z لمنحنى التوزيع الطبيعي القياسي N(0,1) شكل (٤ , ٨) كما يلي :



شكل (٤ , ٨)

ويتضح من منحنى التوزيع الطبيعي القياسي شكل (٤ , ٨) أن غالبية قيم z تقع داخل الفترة (3,3-) أو بمعنى آخر فإنه نادر ما نجد قيمة z خارج هذه الفترة . ويمكن

إيجاد قيمة الاحتمـال z داخل أي فترة باستخدام العلاقة (٤,٢٦) حيث أمكن عمل جداول إحصائية تعطي احتمال وقوع المتغير Z داخل أي فترة . جدول (١) (آخر الكتاب) . حيث إن احتمال وقوع Z بين الصفر وأية قيمة موجبة z تعطى بالاحتمال $P(0 \quad Z \quad z)$ ويمثل بالمساحة المظللة في شكل (٤,٩) كما سيـلـي . ويمكن الحصول على هذه المساحة (الاحتمال) من جدول (١) (آخر الكتاب) وتبدأ z من الصفر وتزداد بمقدار 0.01 ونظرا لأن المنحنى متماثل حول $z = 0$ فإنه يمكن استخدام جـدول (1) لايجاد المساحة (الاحتمال) المحصورة بين الصفر وقـيـم z السالبة مثل ما يتم بين الصفر وقيم z الموجبة وذلك باستخدام التماثل حول محور التناظر $z = 0$ ونوضح ذلك بالأمثلة كما يلي.

مثال (٤,١٤)

إذا علم أن أطوال مجموعة من الطلاب يتبع التوزيع الطبيعي $N(165,25)$

ا) أوجد القيم المعيارية z إذا كان أطوال الطلاب X تأخذ القيم 150,165,172 .

ب) أوجد الأطوال الحقيقية X إذا كـانت القيم المعيـارية للطـلاب 0.52, 0, 1.2 - .

الحل

القيم المعيارية z تعطى بالعلاقة $Z = \dfrac{X - \mu}{\sigma}$ حيث $\mu = 165, \sigma = 5$.

فإن Z_1 القيمة المعيارية للطالب الأول حيث $x_1 = 150$ هي :

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{150 - 165}{5} = -3$$

وأن Z_2 حيث $x_2 = 165$ للطالب الثاني هي :

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{165 - 165}{5} = 0$$

وأن Z_3 حيث $x_3 = 172$ للطالب الثالث هي :

$$z_3 = \frac{x_3 - \mu}{\sigma} = \frac{172 - 165}{5} = 1.4$$

وبالمثل تحسب القيم الحقيقية x لأطوال الطلاب من العلاقة

$$x = \mu + z\sigma$$

فإن x_1 طول الطالب الأول الحقيقي حيث $z_1 = -0.52$ هو :

$$x_1 = 165 + (-0.52)\,5 = 162.4 \quad \text{سم}$$

وإن x_2 طول الطالب الثاني الحقيقي حيث $z_2 = 0$ هو :

$$x_2 = 165 + (0)\,5 = 165 \quad \text{سم}$$

وأن x_3 طول الطالب الثالث الحقيقي حيث $z_3 = 1.2$ هو :

$$x_3 = 165 + (1.2)\,5 = 171 \quad \text{سم}$$

مثال (٤ , ١٥)

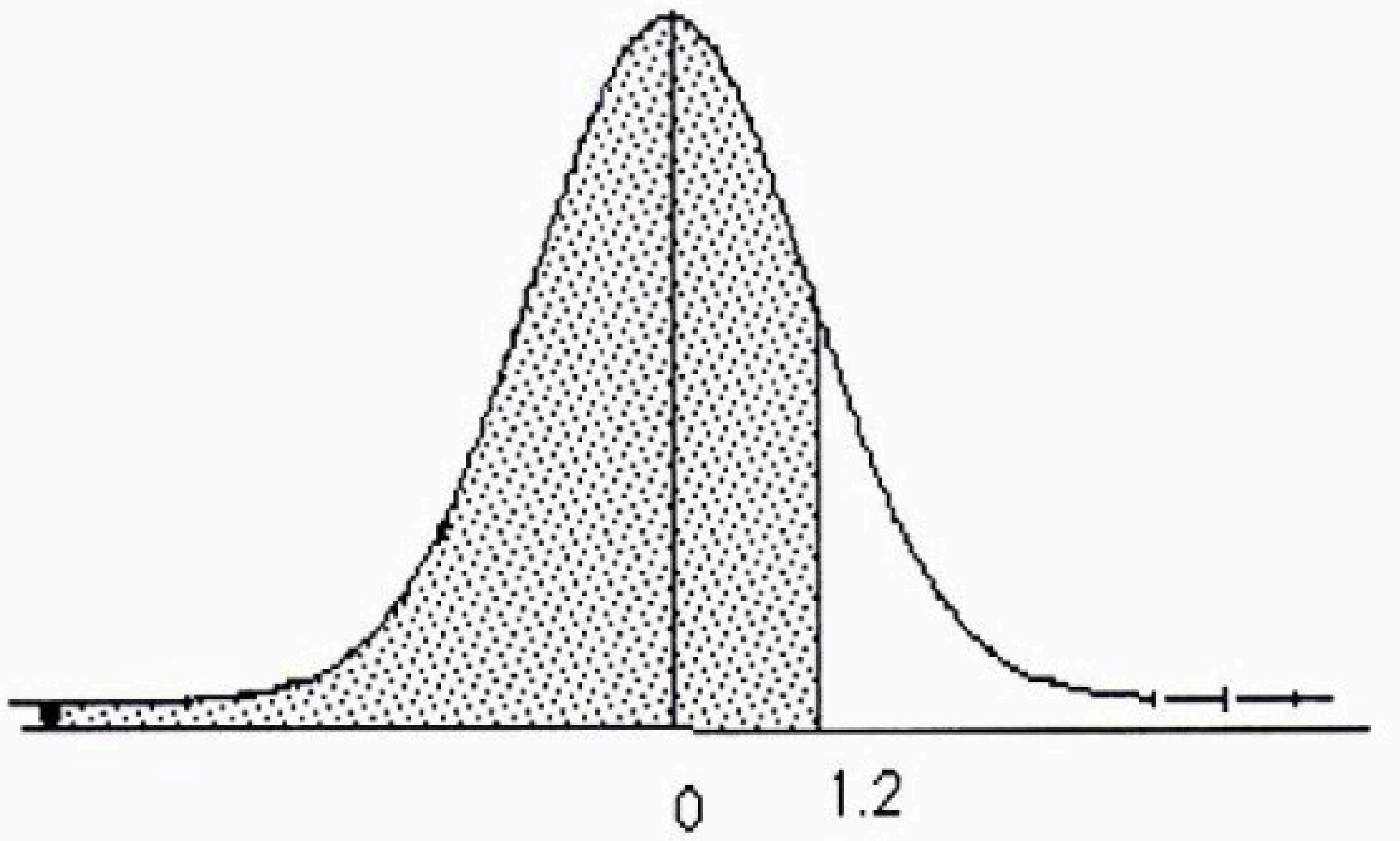
المتغير العشوائي Z موزع توزيعا طبيعيا قياسيا $N(0,1)$ باستخدام الجدول الإحصائي (جدول (١)) أوجد الاحتمالات التالية :

(i) $P(Z \le 1.2)$, (ii) $P(Z \le -0.54)$, (iii) $P(Z \ge 0.25)$

(iv) $P(-1.31 \le Z \le 0.45)$.

الحل

(i) نوضح الاحتمال $P(Z \le 1.2)$ بالمساحة المظللة تحت المنحنى للتوزيع الطبيعي القياسي $N(0,1)$ شكل (٤ , ٩) كما يلي :



شكل (٩ , ٤)

وحيث إن المساحة على جانبي Z = 0 تساوي 0.5 فإن المساحة المظللة شكل (٩ , ٤) تمثل الاحتمال المطلوب حيث

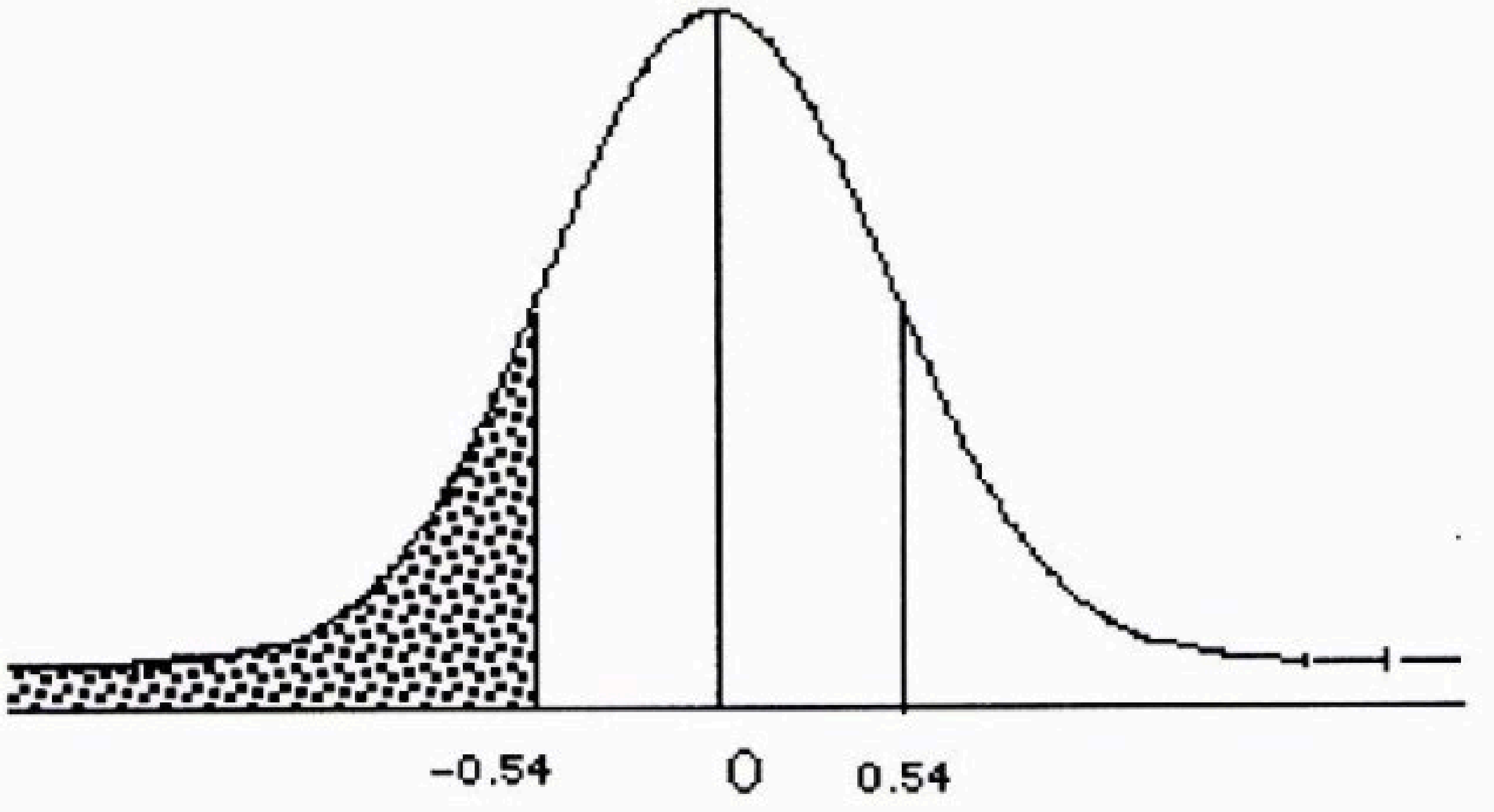
$$P(Z \leq 1.2) = 0.5 + P(0 \leq Z \leq 1.2)$$

ونقرأ الاحتمـــال $P(0 \leq Z \leq 1.2)$ من الجدول (١) وهي القيمة التي بالجدول أمام Z = 1.2 واسفل الرأسي (00) كما يلي :

$$P(Z \leq 1.2) = 0.5 + 0.3849 \quad = 0.8849$$

أي أن الاحتمال المطلوب يمثل نسبة 88% من المساحة تحت المنحنى شكـل (٩ , ٤) السابق والتي مساحته الكلية واحدا صحيحا .

(ii) وبالمثل يوضح الاحتمـال $P(Z \leq -0.54)$ بالمساحة المظللة تحت المنحنـى شكل (١٠ , ٤) كما يلي :



شكل (١٠ , ٤)

وحيث إن المساحة للفترة $(-\infty, -0.54)$ مساوية للمساحة للفترة $(0.54, \infty)$ من التناظر حول $Z = 0$ في الشكل (١٠ , ٤) السابق فإن الاحتمال المطلوب يحسب كما يلي :

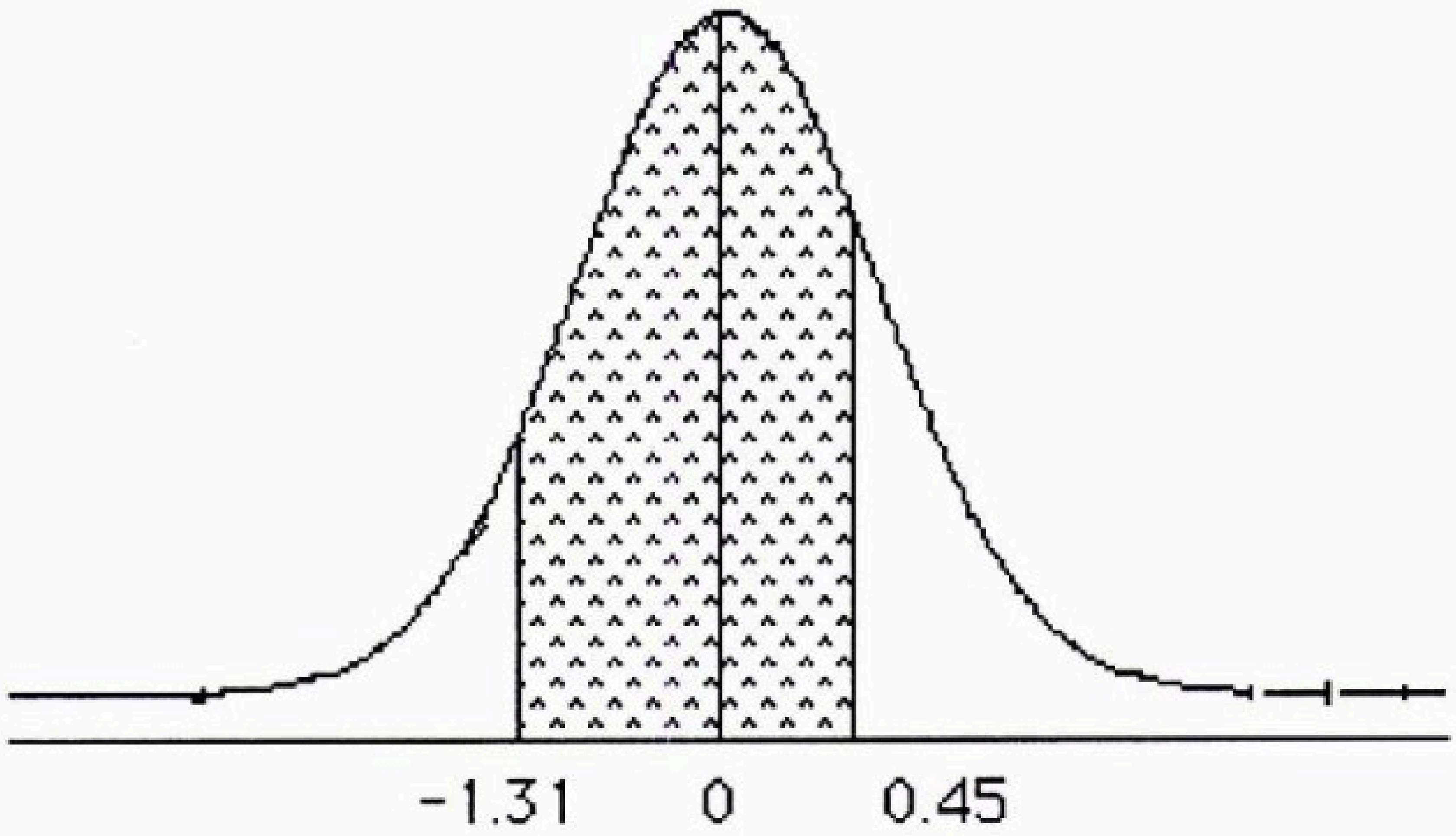
$$P(Z < -0.54) = 0.5 - P(0 \leq Z \leq 0.54)$$

وتقرأ قيمة $P(0 \leq Z \leq 0.54)$ من الجدول (١) آخر الكتاب أمام $Z = 0.5$ وأسفل الرأسي (04) فإن الاحتمال المطلوب يحسب كما يلي:

$$P(Z \leq -0.54) = 0.5 - 0.2054 = 0.2946$$

أي أن الاحتمال المطلوب يمثل نسبة 29% من مساحة المنحنى شكل (١٠ , ٤) .

(iii) وبالمثل يوضح الاحتمال $P(-1.31 \leq Z \leq 0.45)$ بالجزء المظلل تحت المنحنى شكل (١١ , ٤) كما يلي:



شكل (١١ , ٤)

المساحة المظللة في شكل (١١ , ٤) تمثل الفترة (1.31, 0.54-) وهي عبارة عن مجموع الفترتين (1.31, 0-) و (0, 0.45) وباستخدام التناظر حول $Z = 0$ فإن المساحة المظللة مكافئة لمجموع الفترتين (0, 131) و (0,0.45) ويكتب الاحتمال المطلوب كما يلي :

$$P(-1.31 \leq Z \leq 0.45) = P(0 \leq Z \leq 1.31) + P(0 \leq Z \leq 0.45)$$

وباستخدام جدول (١) آخر الكتاب نحصل الاحتمال المطلوب كما يلي:

$$P(-1.31 \leq Z \leq 0.45) = 0.4049 + 0.1736 = 0.5785$$

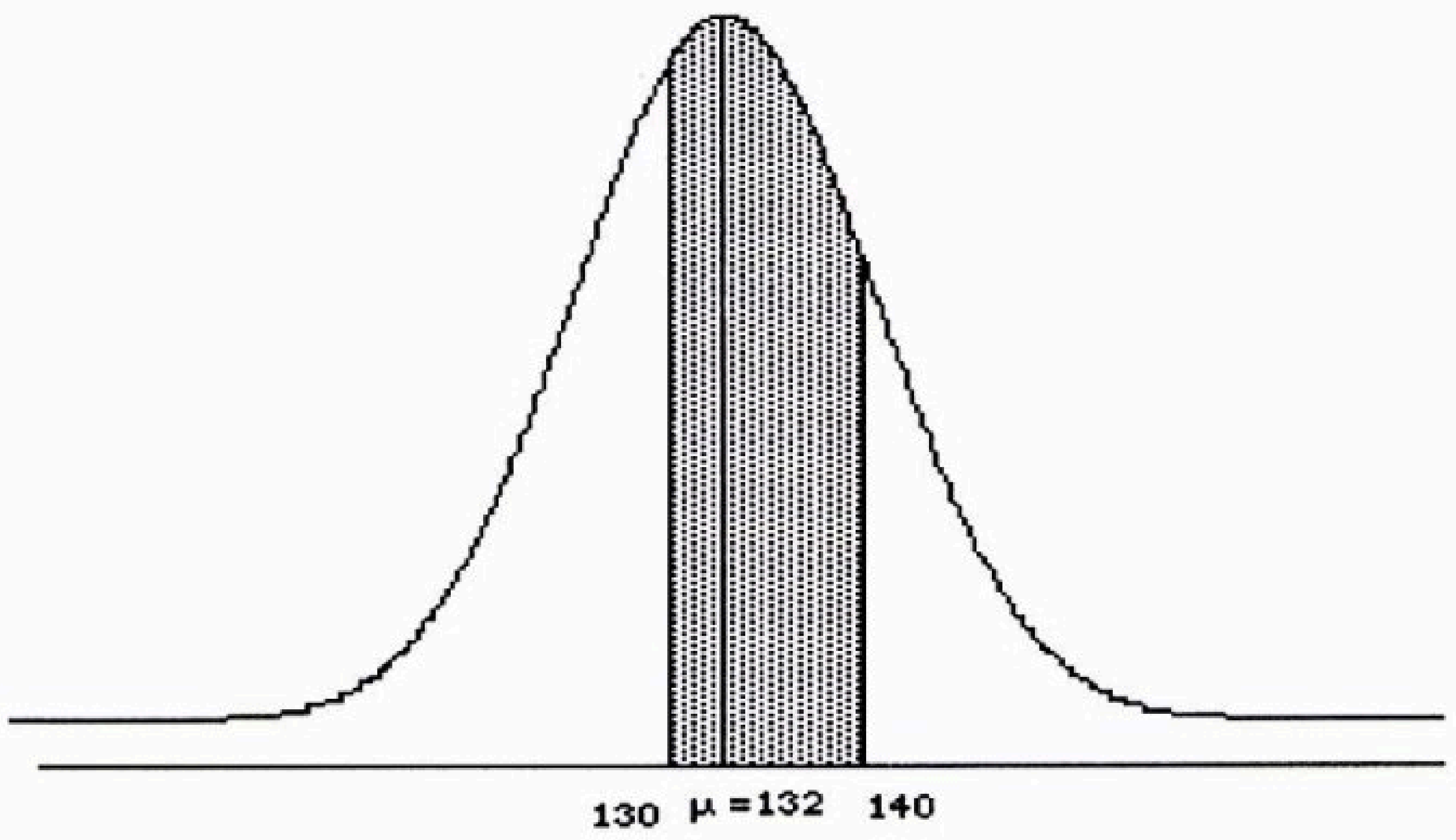
أي أن الاحتمال المطلوب يمثل نسبة 58% من مساحة المنحنى شكل (١١ , ٤) السابق.

مثال (١٦ , ٤)

إن كانت أطوال 500 ورقة من أوراق نبات من مجتمع طبيعي N(132, 100) مختلفة. أوجد عدد الأوراق ما بين 130 ملم و 140 ملم .

الحل

نوضح الاحتمال المطلوب $P(130 \le X \le 140)$ بالجزء المظلل في شكل (٤,١٢) للتوزيع الطبيعي الذي متوسطه 132 وانحرافه المعياري 10 كما يلي:



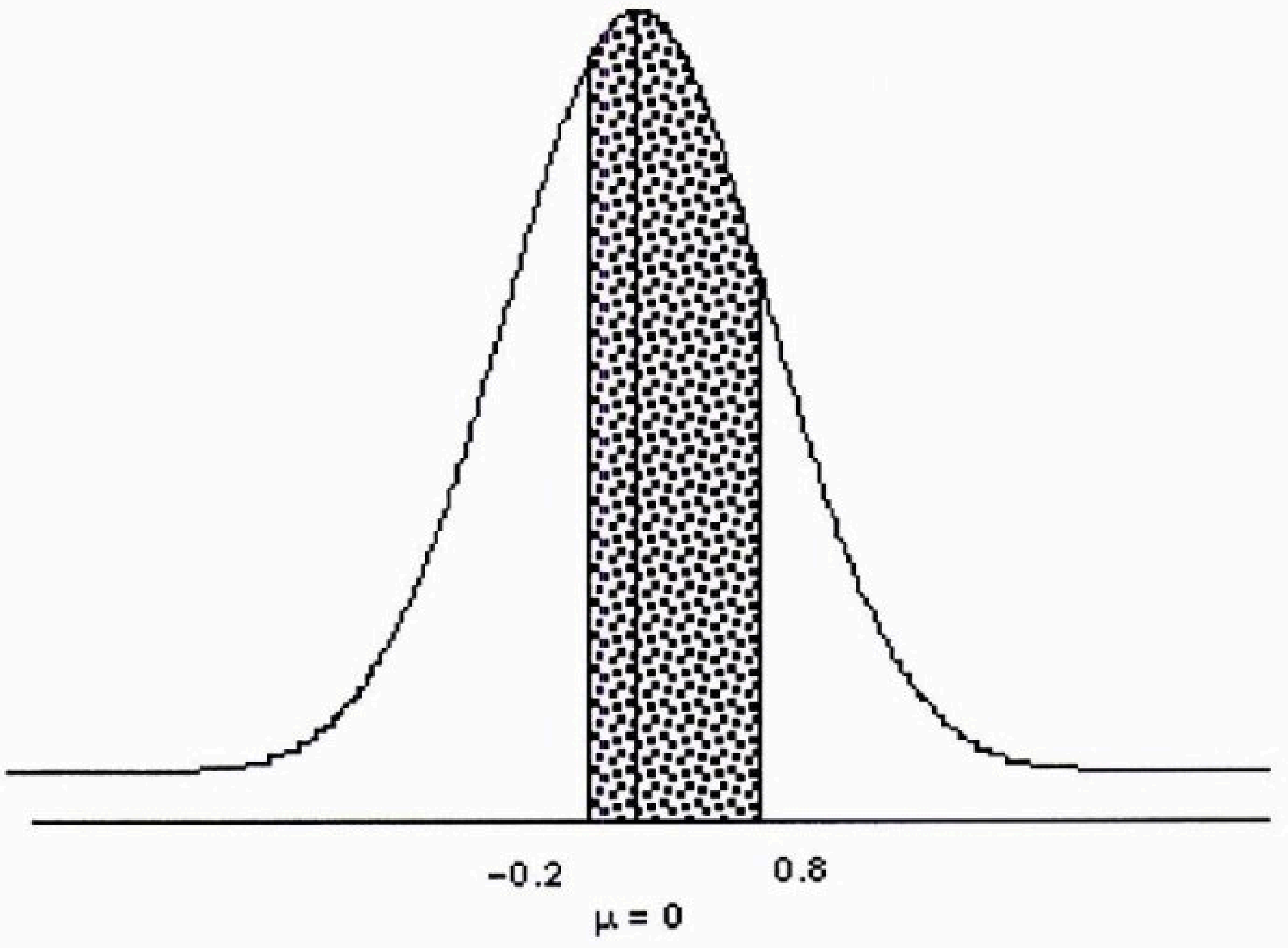
شكل (٤,١٢)

نحول المتغير X = طول الورقة بالمليمتر إلى المتغير القياسي Z حيث $\mu = 132$ ملم والانحراف المعياري $\sigma = 10$ ملم حيث $Z = \dfrac{X - \mu}{\sigma}$ ويكون الاحتمال المطلوب مكافئا للاحتمال التالي:

$$P\{ \frac{(130 - 132)}{10} \le Z \le \frac{(140 - 132)}{10} \}$$

$$= P(-0.2 \le Z \le 0.8)$$

ويمثل الاحتمال $P(-0.2 \le Z \le 0.8)$ بالمساحة المظللة تحت منحنى التوزيع الطبيعي القياسي $N(0,1)$ شكل (٤,١٣) كما يلي:



شكل (١٣ , ٤)

والاحتمال المطلوب يمثل المنطقة المظللة على الـفـتـرة (0.8 ,0.2-) ويكافئ مجموع الاحتمال على الفترتين (0,0,8) و (0,0.2) التي يمكن حسابهما من جدول (١) آخر الكتاب كما يلي :

$$P(-0.2 \le Z \le 0.8) = P(0 \le Z \le 0.2) + P(0 \le Z \le 0.8)$$
$$=0.0793+0.2881=0.3674$$

أي أن نسبة الأوراق داخل الفـتـرة (130, 140) هي تقريبـا 37% . ويحسب عـدد الأوراق داخل هذه الفترة كما يلي :

عدد الأوراق في المجموعة = $500 \times \frac{37}{100} = 185$ ورقة .

(٣, ٦, ٤) تقريب ذي الحدين بالتوزيع الطبيعي

(Normal approximation to the binomial)

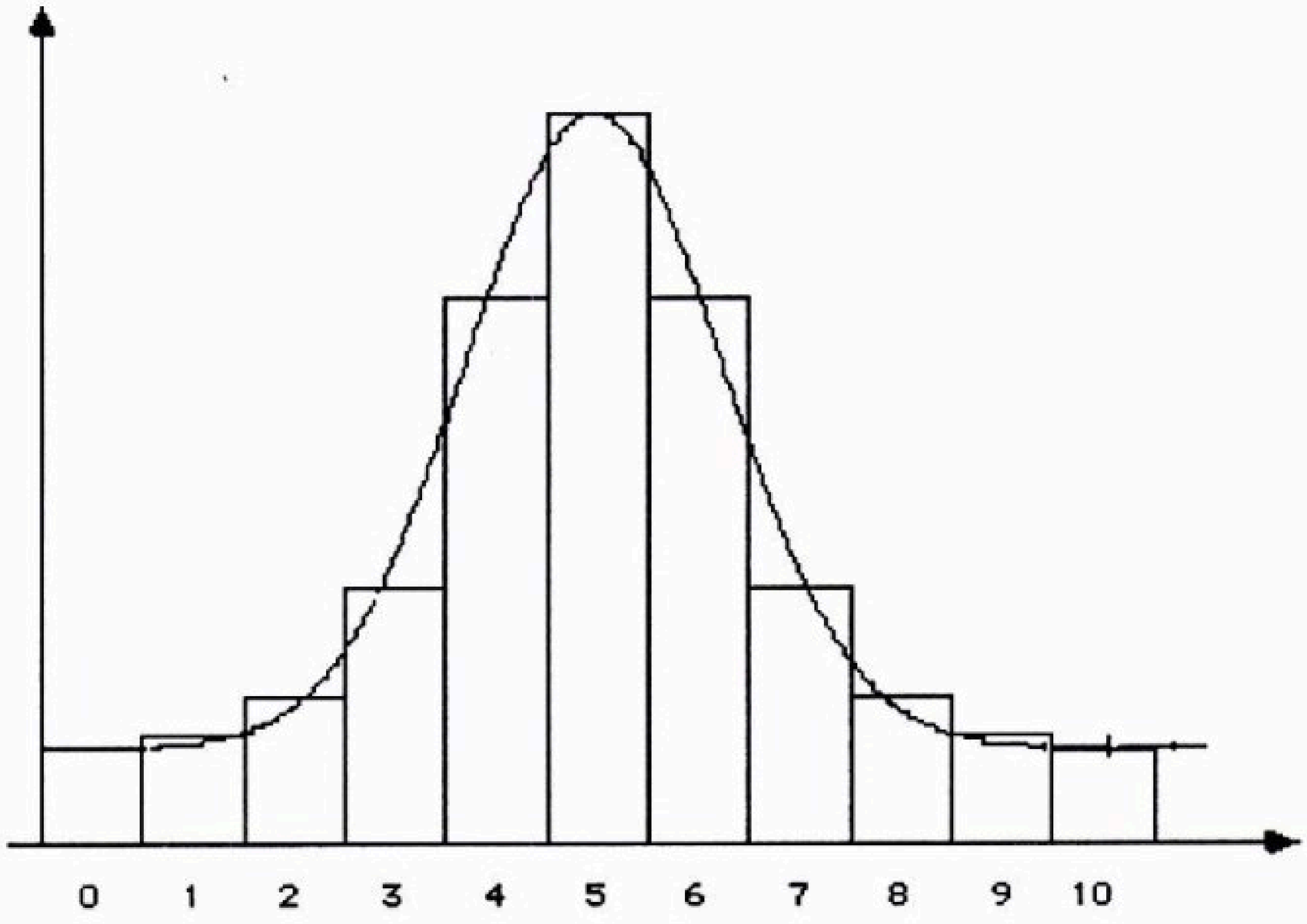
تحسب الاحتمالات للمتغير العشوائي X الذي يتبع توزيع ذي الحدين من العلاقة التالية :

$$b(x, n, p) = P(X = x) = \binom{n}{x} p^{x} q^{n-x} \qquad (٤, ٢٧)$$

ويستخدم التوزيع الطبيعي (وهو توزيع متصل) لتقريب الاحتمال لتوزيع ذي الحدين (وهو توزيع متقطع) والمعطى بالعلاقة (٤, ٢٧) ، وذلك عندما تكون n عدد المحاولات كبيرا $(n > 20)$ والاحتمال p (احتمال النجاح) قريب من 0.5 ، أما التوزيع الطبيعي الذي يستخدم في التقريب يجب أن يكون متوسطه وتباينه هو نفس متوسط وتباين توزيع ذي الحدين المطلوب حساب الاحتمالات له . أي أن $\mu = n p,\ \sigma^{2} = n p q$ كما في شكل (٤, ١٤) يمثل المدرج التكراري والمنحنى التكراري لمتغير X الذي له توزيع ذي الحدين وناتج من تجربة عشوائية هي قذف قطعة عملة متزنة عشر مرات حيث X = عدد الصور فإن القيم الممكنة للمتغير X هو المجموعة X(s) كما يلي :

$$X(s) = \{0, 1, 2, ..., 10\}$$

ويرسم المدرج التكراري النسبي للقيم الممكنة للمتغير X حيث إن قيمة المتغير $x = 6$ هي منتصف قاعدة المستطيل على المحور الأفقي وقيمة الاحتمال $P(X = 6)$ يمثل ارتفاع هذا المستطيل وهكذا بالنسبة لجميع القيم الممكنة فنحصل على المدرج التكراري النسبي وكذلك المنحنى التكراري النسبي كما هو مبين في شكل (٤, ١٤) كما يلي:



شكل (٤,١٤)

ويحسب الاحتمال $P(X = 6)$ من العلاقة (٤,٢٧) كما يلي :

$$P(X = 6) = b\left(6,\ 10,\ \frac{1}{2}\right) = \binom{10}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^4 = 0.205$$

$$\mu = np = 10\left(\tfrac{1}{2}\right) = 5$$

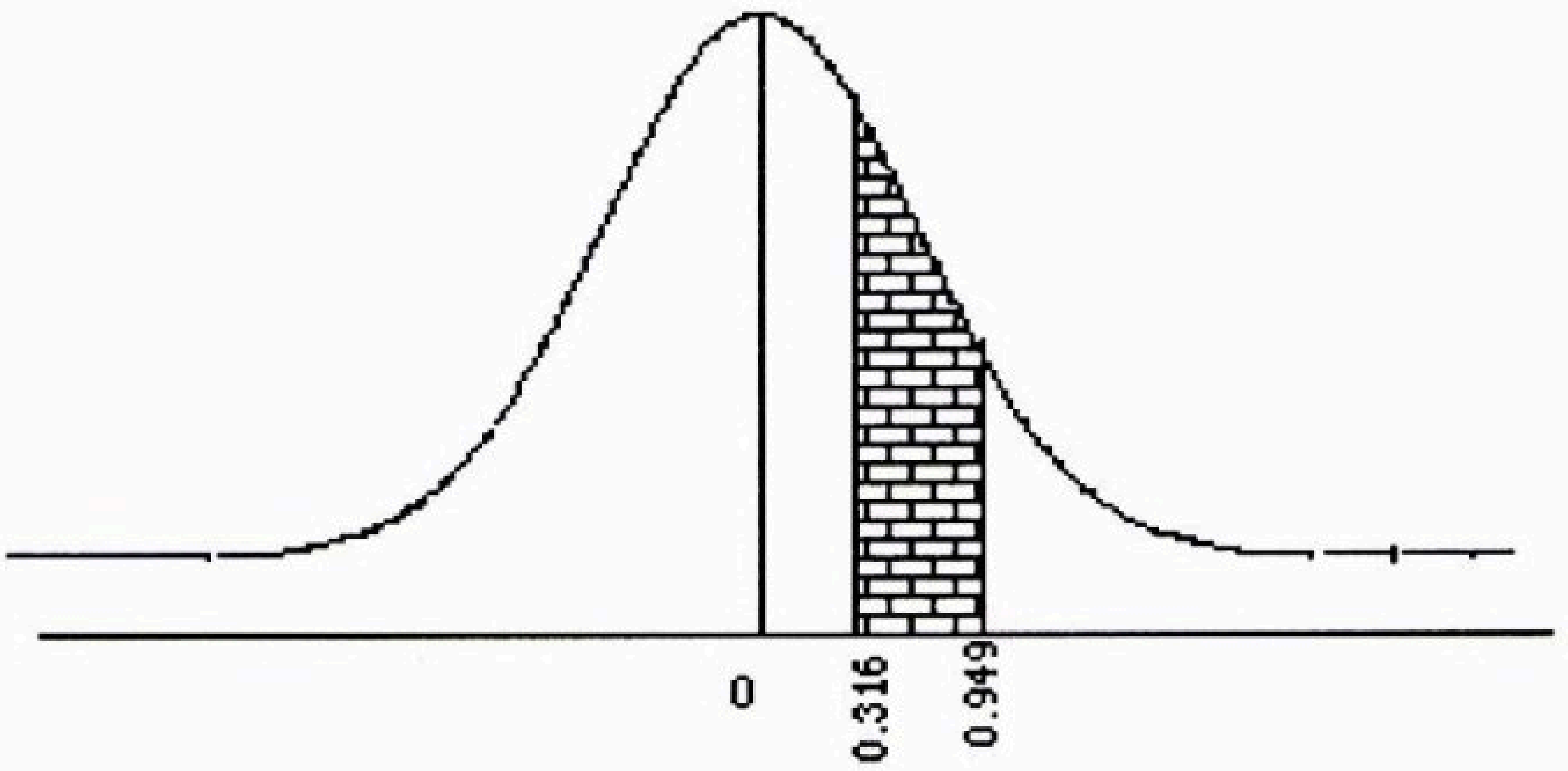
$$\sigma = \sqrt{npq} = \sqrt{10\left(\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right)} = 1.58$$

وهذا الاحتمال $P(X = 6)$ يمثل المساحة للمستطيل الذي قاعدته الـفـتـرة $(5.5, 6.5)$ وتقرب مساحة المستطيل بواسطة التوزيع الطبيعي القياسي $N(0,1)$ كما يلي :

$$P(5.5 \le X \le 6.5) = P\left(\frac{5.5 - 5}{1.58} \le Z \le \frac{6.5 - 5}{1.58}\right)$$

$$= P(0.316 \le Z \le 0.949)$$

ونوضح الاحتمال المطلوب بالجزء المظلل في الشكل (٤,١٥) كما يلي:



شكل (٤,١٥)

ومن جدول (١) آخر الكتاب نحصل على قيمة الاحتمال للجزء المظلل كما يلي:

$$P(0.316 \le Z \le 0.949) = P(0 \le Z \le 0.949) - P(0 \le Z \le 0.319)$$

$$= 0.3289 - 0.1253 = 0.2034$$

ويظهر الفرق بين الإجابتين أقل من 0.002 . وهذا الفرق ناتج عن التقريب وممكن يكون أقل من ذلك كثيرا كلما زادت قيمة n .

مثال (٤ , ١٧)

إذا القيت قطعة عملة متزنة مائة مرة وكان المتغير العشوائي X = عدد مرات ظهور الصورة . استخدم التقريب إلى التوزيع الطبيعي لحساب احتمال ظهور الصورة عدد 57 مرة على الأكثر ($P(X \le 57)$) .

الحل

إذا استخدمنا توزيع ذي الحدين $b(x, n, p) = b(57, 100, \frac{1}{2})$ فإن

الاحتمال $P(X \le 57)$ هو :

$$P(X \le 57) = \binom{100}{0}(\frac{1}{2})^{0}(\frac{1}{2})^{100} + \binom{100}{1}(\frac{1}{2})^{1}(\frac{1}{2})^{99} + \ldots + \binom{100}{57}(\frac{1}{2})^{57}(\frac{1}{2})^{43}$$

أي أن المطلوب حساب 58 حدا وهذا يتطلب جهدا كبيرا .

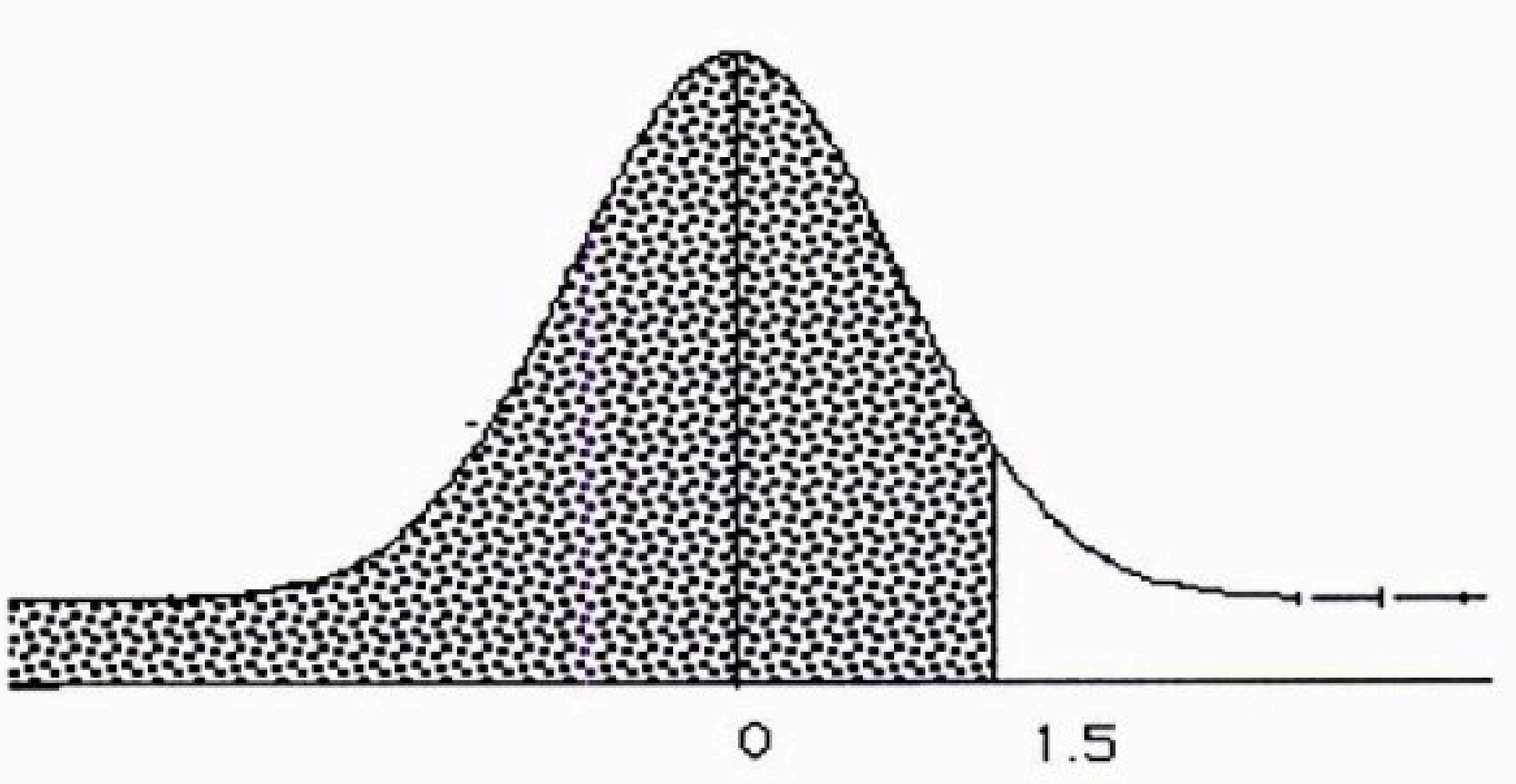
ولكن باستخدام التقريب الطبيعي يتم الحساب بسهولة كما يلي :

$$\mu = n\,p = 100(\frac{1}{2}) = 50, \quad \sigma = \sqrt{100(\frac{1}{2})(\frac{1}{2})} = 5$$

والاحتمال المطلوب كما يلي :

$$P(X \le 57.5) = P(Z \le \frac{57.5 - 50}{5}) = P(Z \le 1.5)$$

ونوضح الاحتمال المطلوب بالجزء المظلل في منحنى التوزيع الطبيعي $N(0,1)$ في شكل (١٦ , ٤) كما يلي :



شكل (١٦ , ٤)

أي أن

$$P(Z \le 57.5) = 0.5 + P(0 \le Z \le 1.5)$$

ومن الجدول (١) آخر الكتاب نجد أن :

$$P(Z \leq 57.5) = 0.5 + 0.43332 = 0.9332$$

أي أن الاحتمال المطلوب يمثل نسبة 93% من مساحة المنحنى شكل (٤ , ١٦) السابق .

ملاحظة :

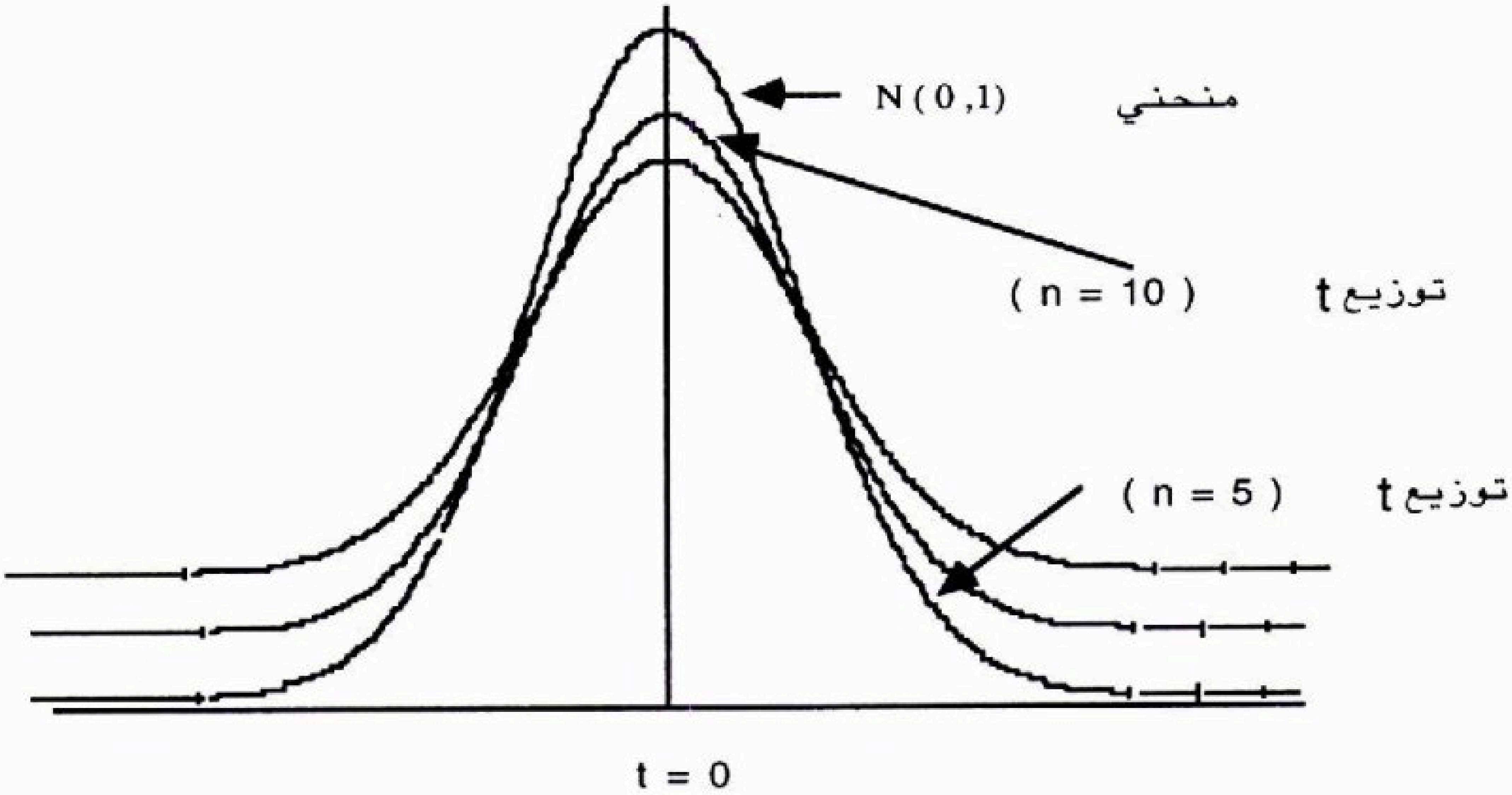
القيمة $X = 57.5$ تمثل نهاية الفترة $(56.5, 57.5)$ التي تمثل قاعدة المستطيل الذي مركزه $X = 57$ والتقريب للطبيعي يتطلب احتواء جميع الفترة $(56.5, 57.5)$ ضمن حساب الاحتمال المطلوب أي أن $P(X \leq 57.5)$.

(٤ , ٦ , ٤) توزيع t (t-Distribution)

إذا كان T متغيرا عشوائيا متصلا دالة كثافته الاحتمالية $f(t)$ كما يلي :

$$f(t) = c\left(1 + \frac{t^2}{\upsilon}\right)^{-\upsilon + \frac{1}{2}} \;,\; -\infty < t < \infty \qquad (٤ , ٢٨)$$

حيث υ تسمى درجات حرية $(\upsilon = n - 1)$ ، c تحدد بالمقدار υ . ولقد اكتشف هذا التوزيع العالم الإيرلندي جوست W.S. Goset عام ١٩٠٨م ونشره تحت اسم مستعار (توزيع t) أو توزيع طالب وهذا التوزيع يشبه التوزيع الطبيعي القياسي ولكن قمته أكثر انخفاضا عن قمة التوزيع الطبيعي القياسي ويتقارب طرفيه من الصفر عند $-\infty, +\infty$ أبطأ من تقارب طرفي التوزيع الطبيعي القياسي ويقترب من التوزيع الطبيعي القياسي كلما كبر حجم العينة n أي كلما كبرت درجات الحرية $\upsilon = n - 1$ وينطبق على التوزيع الطبيعي القياسي عندما $\upsilon = 30$ ونوضح ذلك في شكل (٤ , ١٧) كما يلي :



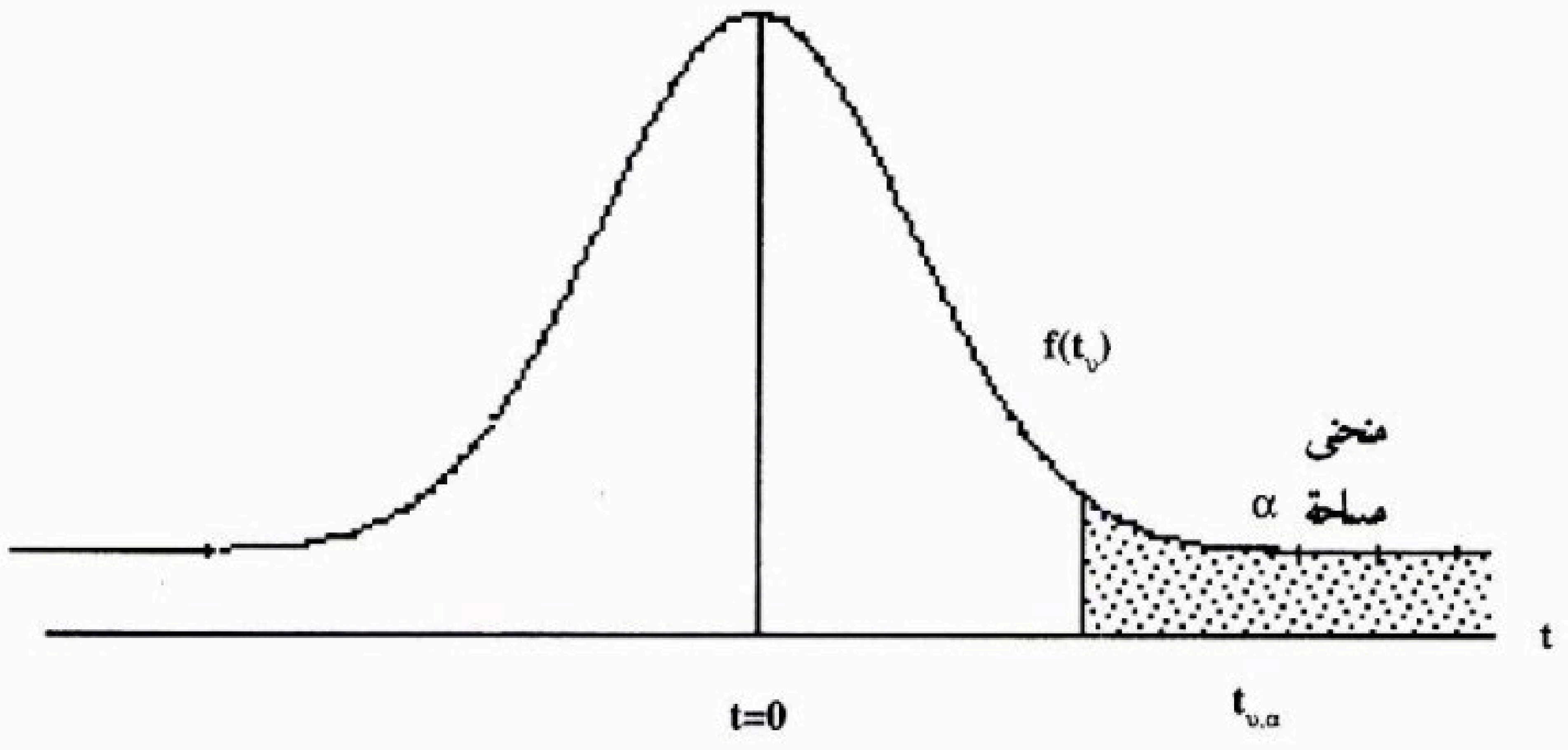
شكل (٤,١٧)

وتوزيع t توقعه صفرا أي أن $\mu = E(T) = 0$ وهو متماثل حول المحور العمودي عند $t = 0$ والمار بقمته وتباينــه $\sigma^2 = V(T) = \dfrac{\upsilon}{\upsilon - 2}$ وهو أكبر من الواحد الصحيــح دائما . كلما زادت درجات الحرية υ اقترب التباين من الواحد الصحيح واقترب من التوزيع الطبيعي القياسي N(0,1) .

لحساب أي احتمالات حول المتغيــر T يلزمنا وجود جدول يبين المســاحــات المختلفة تحت منحنى الدالة $f(t_\upsilon)$ لهذا المتغير والمحصورة بين قيمتين من قيم المتغير كما هو الحال في حالة التوزيع الطبيعي القياسي . ولكن يوجد لكل درجة حرية υ منحنى مختلف . وهذا يعني أنه يلزمنا جدول خاص بكل قيمة من قيــم υ وهذه عملية شاقة . وحيث إن الاستخدامات الإحصائية لتوزيع t تعتمد على معرفة قيم t التي وجد على يمينها مساحة صغيرة α ، فإن المطلوب هو معرفة القيمة $t_{\upsilon,\alpha}$ حيث يحسب الاحتمال كما يلي :

$$P(t_{\upsilon} \geq t_{\upsilon,\alpha}) = \alpha \qquad (٤ , ٢٩)$$

والمساحة المظللة α على يمين $t_{\upsilon,\alpha}$ في شكل (٤ , ١٨) توضح كما يلي :



شكل (٤ , ١٨)

ومن الجدول (٢) في آخر الكتاب تحسب قيم $t_{\upsilon,\alpha}$ عند قيم خاصة من α توضح في الجدول (٢) عند القيم الشائعة الاستخدام وهي

$$\alpha = 0.005, 0.01, 0.025, 0.05, 0.10$$

وذلك لكل درجة حرية υ التي تأخذ قيم من 1 إلى 30 درجة .

ملاحظة

حيث إن توزيع t متناظر حول $t = 0$ فإن $t_{\upsilon,1-\alpha}$ تساوي قيم $-t_{\upsilon,\alpha}$.

مثال (٤ , ١٨)

إذا كانت $\upsilon = 15$ فأوجد قيم $t_{\upsilon,\alpha}$ التي تجعل الاحتمالات التالية :

(i) $P(t_{\upsilon} \geq t_{\upsilon,\alpha}) = 0.005$

(ii) $P(t_{\upsilon} \geq t_{\upsilon,\alpha}) = 0.01$

(iii) $P(t_{\upsilon} \geq t_{\upsilon,\alpha}) = 0.025$

(iv) $P(t_{\upsilon,0.99} \leq t_{\upsilon} \leq t_{\upsilon,0.05})$

الحل

باستخدام العلاقة (٤,٢٩) وبالبحث في جدول (٢) لتوزيع t آخر الكتاب نجد أن :

$$(i) \quad t_{15,0.005} = 2.947$$

والقيمة $t_{15,0.005} = 2.947$ في الجدول هي القيمة النـاتجة من تقاطـع الصـف $\upsilon = 15$ والعمود $\alpha = 0.005$ وبالمثل نوجد القيم الباقية للمثال (٤,١٧) كما يلي:

$$(ii) \quad t_{15,0.01} = 2.602$$

وهي القيمة الناتجة من تقاطع الصف $\upsilon = 15$ بالعمود $\alpha = 0.01$.

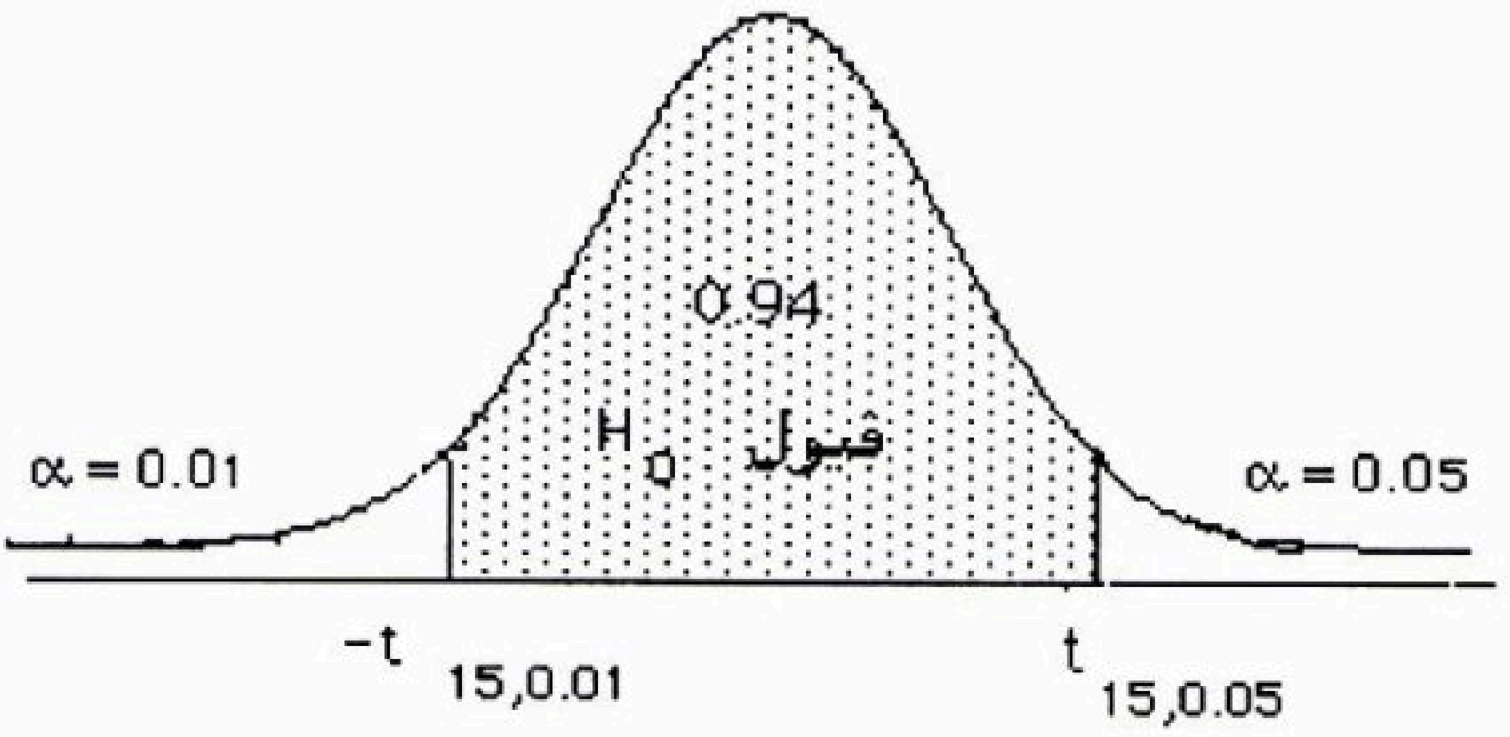
$$(iii) \quad t_{15,0.025} = 2.131$$

وهي القيمة الناتجة من تقاطع الصف $\upsilon = 15$ بالعمود $\alpha = 0.025$.

$$(iv) \quad t_{15,0.99} = t_{15,0.01} = -2.602$$

$$t_{15,0.05} = 1.725$$

فإن الاحتمال المطلوب يمثل المنطقة المظللة في شكل (٤,١٩) كما يلي :



شكل (٤,١٩)

$$P(t_{15,0.99} \le t_{15} \le t_{15,0.05}) = 0.99 - 0.05 = 0.94$$

أي أن

$$P(-2.602 \le t_{15} \le 1.725) = 0.94$$

(٤,٦,٥) توزيع مربع كاي (Chi-square distribution)

إذا كان χ^2 متغيرا عشوائيا دالة كثافته الاحتمالية $f(\chi^2)$ كما يلي :

$$f(\chi^2) = c(\chi^2)^{\frac{\upsilon}{2}-1} e^{-\frac{x^2}{2}} \quad , \quad \chi^2 > 0 \qquad (٤,٣٠)$$

حيث υ تسمى درجات الحرية ، c مقدار يحدد بدرجات الحرية υ . والدالة $f(\chi^2)$ تسمى دالة كثافة توزيع مربع كاي ذات درجات حرية υ حيث المقدار c يحدد بدرجات الحرية بحيث تكون المساحة تحت المنحنى 1 . ويستخدم مربع كاي في كثير من التطبيقات الإحصائية وفيما يلي نلخص بعض خصائص هذا التوزيع.

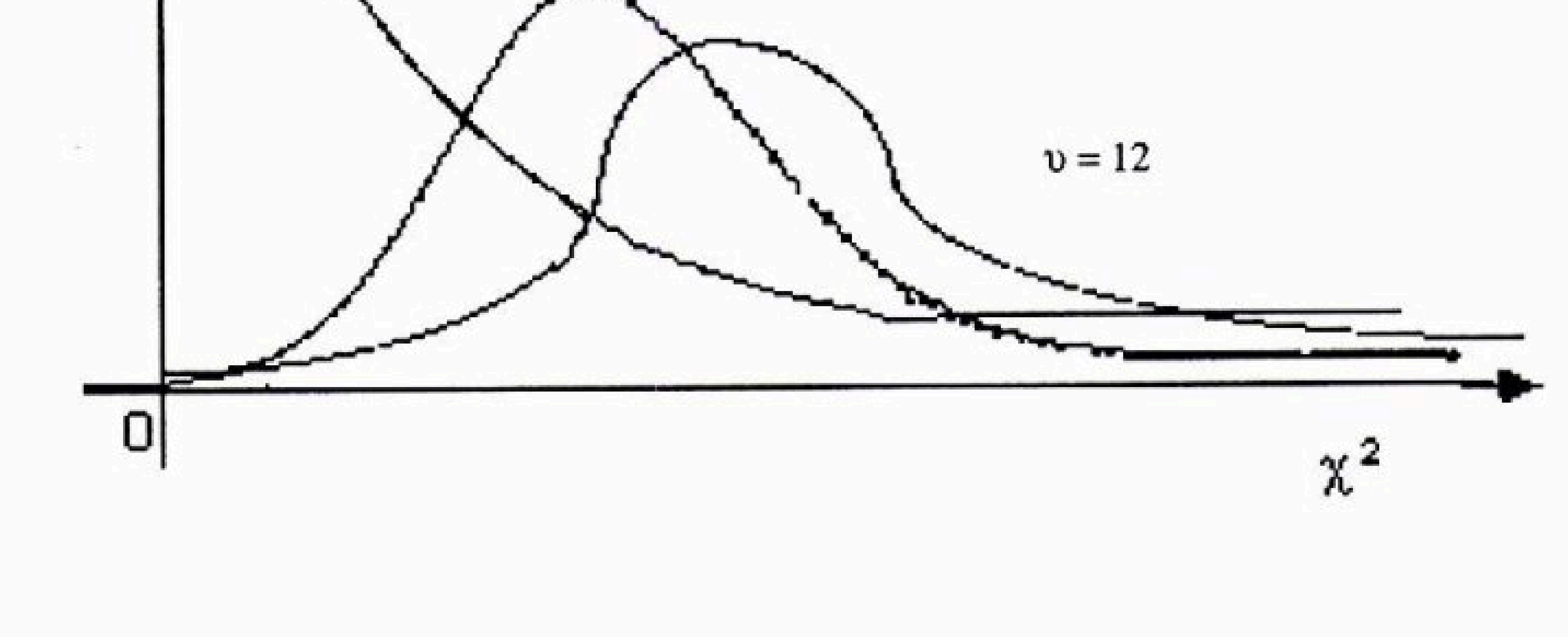
١ - توزيع مربع كاي متصل ، حيث يمكن حساب الاحتمال بالمساحات تحت

المنحنى الذي يعتمد على درجات الحـريـة υ حيث يوجد عدد كبير مـن التوزيعات لمربع كاي تعتمد على تغير درجة الحرية υ .

٢ - قيم المتغير χ^2 موجبة وبالتالي فإن جميع المنحنيات لمربع كاي تقع على يمين المحور الرأسي $\chi^2_\upsilon = 0$ أي أن الفترة $(0, \infty)$.

٣ - تحدد درجات الحرية υ شكل المنحنى لتوزيع مربع كاي وأن متوسطه وتباينه هما $\mu = \upsilon \ , \quad V(\chi^2) = 2\upsilon$. كلما زادت درجات الحرية كلما زاد المتوسط والتباين وكذلك تفلطح منحنى التوزيع لمربع كـاي واقترابـه من التوزيع الطبيعي عند $\upsilon = 30$.

نوضح في شكل (٢٠ , ٤) تغير منحنى توزيع مربع كاي حسب تغيـر درجـات الحرية υ كما يلي :



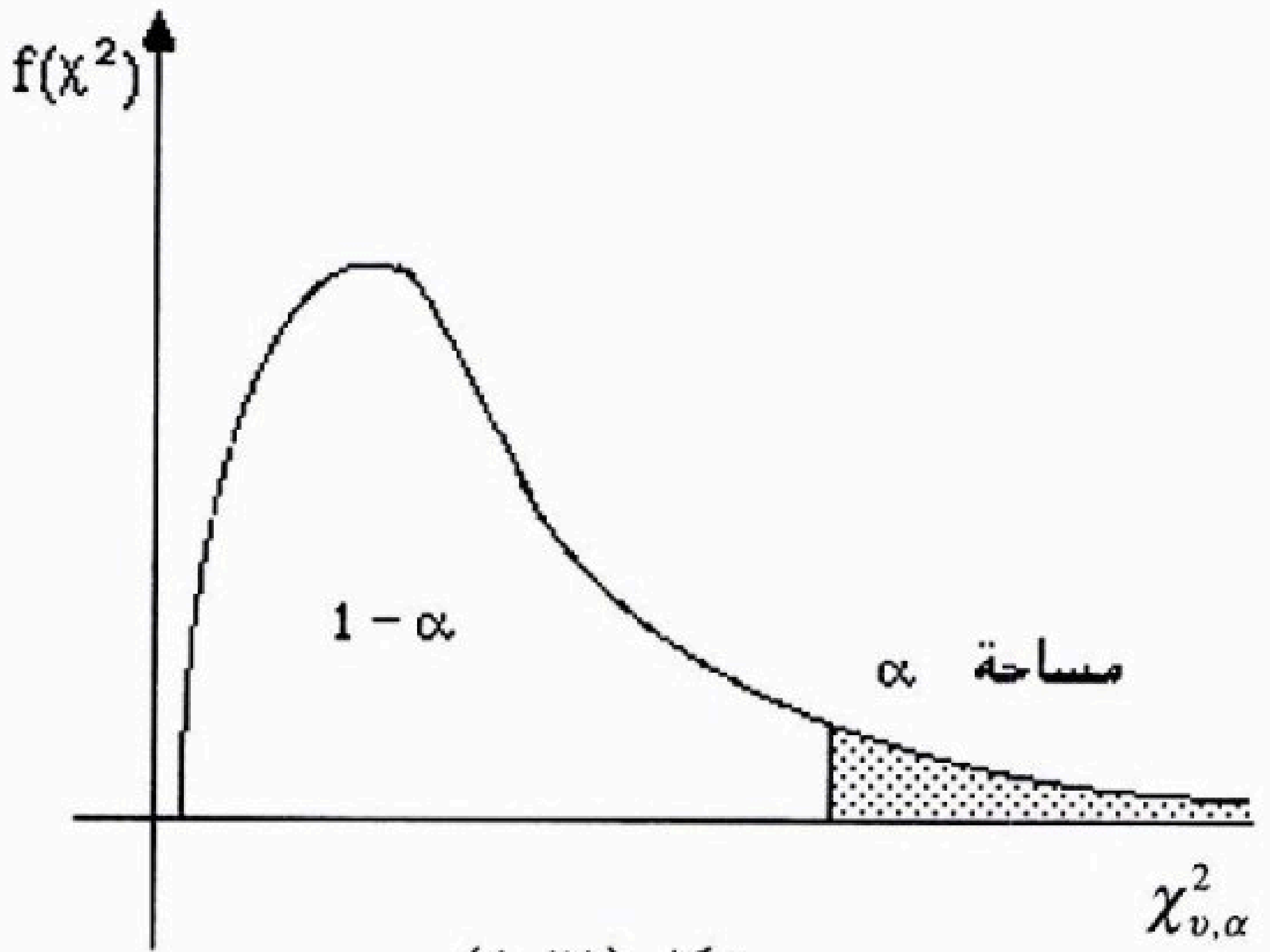
شكل (٢٠ , ٤)

والجدول (٣) آخر الكتاب يوضح قيم $\chi^2_{\upsilon,\alpha}$ التي تكون على يمين هذه القيمة المظلله α كما هو موضح في شكل (٤,٢١) . تكتب بعض قيم α ذات الأهمية في التطبيقات الإحصائية مثل :

$$\alpha = 0.995, 0.99, 0.975, 0.95, 0.05, 0.025, 0.01, 0.005$$

ويتم معرفة القيمة $\chi^2_{\upsilon,\alpha}$ من الجدول (٣) آخر الكتاب لقيم α المعطاة أعلاه ، حيث الاحتمال α بالعلاقة (٤,٣١) أو نسبة الجزء المظلل على يمين $\chi^2_{\upsilon,\alpha}$ الشكل (٤,٢١) كما يلي :

$$P(\chi^2_{\upsilon} \geq \chi^2_{\upsilon,\alpha}) = \alpha \qquad (٤,٣١)$$



شكل (٤,٢١)

مثال (٤,١٨)

بالاستعانة بالعلاقة (٤,٣١) والمساحة المظللة في الشكل (٤,٢١). أوجد قيم $\chi^2_{\upsilon,\alpha}$ ، حيث $\upsilon = 10$ للاحتمالات التالية :

$$\text{(i)} \quad P(\chi^2_{\upsilon} \geq \chi^2_{\upsilon,\alpha}) = 0.005$$

$$\text{(ii)} \quad P(\chi^2_{\upsilon} \geq \chi^2_{\upsilon,\alpha}) = 0.01$$

$$\text{(iii)} \quad P(\chi^2_{\upsilon} \geq \chi^2_{\upsilon,\alpha}) = 0.025$$

$$\text{(iv)} \quad P(\chi^2_{\upsilon,\,0.99} \leq \chi^2_{\upsilon} \leq \chi^2_{\upsilon,\,0.05})$$

الحل

(i) وحيث $\alpha = 0.005$ ، $\upsilon = 10$ فإن

$$\chi^2_{\upsilon,\alpha} = \chi^2_{10,\,0.005} = 25.198$$

25.198 هي القيمة الناتجة من التقاء الصف $\upsilon = 10$ والعمود 0.05

(ii) وحيث $\alpha = 0.01$ ، $\upsilon = 10$ فإن

$$\chi^2_{\upsilon,\alpha} = \chi^2_{10,\,0.01} = 23.209$$

23.209 هي القيمة الناتجة من التقاء الصف $\upsilon = 10$ والعمود 0.01 .

(iii) وحيث $\alpha = 0.025$ ، $\upsilon = 10$ فإن

$$\chi^2_{\upsilon,\alpha} = \chi^2_{10,\,0.025} = 20.483$$

20.483 هي القيمة الناتجة من التقاء الصف $\upsilon = 10$ والعمود 0.025

(iv) $$P(\chi^2_{10,\,0.99} \leq \chi^2_{10} \leq \chi^2_{10,\,0.05}) = 0.99 - 0.05 = 0.94$$

من جدول (٣) آخر الكتاب نوجد القيمتين وهما :

$$\chi^2_{10,\,0.99} = 2.508$$

$$\chi^2_{10,\,0.05} = 18.307$$

فيكون الاحتمال المطلوب هو :

$$P(2.508 \le \chi^2 \le 18.307) = 0.94$$

(٤,٦,٦) توزيع F (F-Distribution)

إذا كان F متغيرا عشوائيا دالة كثافته الاحتمالية f(F) كما يلي:

$$f(F) = \frac{cF^{\frac{\upsilon_1}{2}-1}}{(\upsilon_2 + \upsilon_1 F)^{-\left(\frac{\upsilon_1}{2}+1\right)}}, F > 0 \qquad (٤,٣٢)$$

فإنه يقال للدالة f(F) بدالة توزيع F بدرجات حرية υ_1, υ_2 ويرمز لقيم المتغير F بالرمز $F(\upsilon_1,\upsilon_2)$ ، c ثابت يعتمد على υ_1, υ_2 ويعين بحيث تصبح للمساحة المنحنى لتوزيع F تساوي واحدا صحيحا. ويوجد عدد كبير من المنحنيات للتوزيع F حسب تغير قيم درجات الحرية (υ_1, υ_2) . وجميع المنحنيات لتوزيع F تقع على يمين المحور الرأسي F=0 أي أن داخل الفترة $(0,\infty)$ حيث إن المتغير $F_{\upsilon_1,\upsilon_2}$ يأخذ قيما موجبه دائما ومنحنى توزيع F ملتو نحو اليمين ويقل الالتواء كلما زادت درجات الحرية (υ_1, υ_2) حتى ينطبق على التوزيع الطبيعي (عند $\upsilon_1 \ge 30, \upsilon_2 \ge 30$). وبالنسبة υ_1, υ_2 فإننا نجد أن υ_2 يظهر في المقام في العلاقة (٤,٣٢) فإنه يعتبر درجات حرية المقام ويعتبر υ_1 درجات حرية البسط ويظهر υ_1 لقيمة المتغير $F_{\upsilon_1,\upsilon_2}$. ومتوسط وتباين توزيع F هما:

$$\mu = \frac{\upsilon_2}{\upsilon_2 - 2}$$

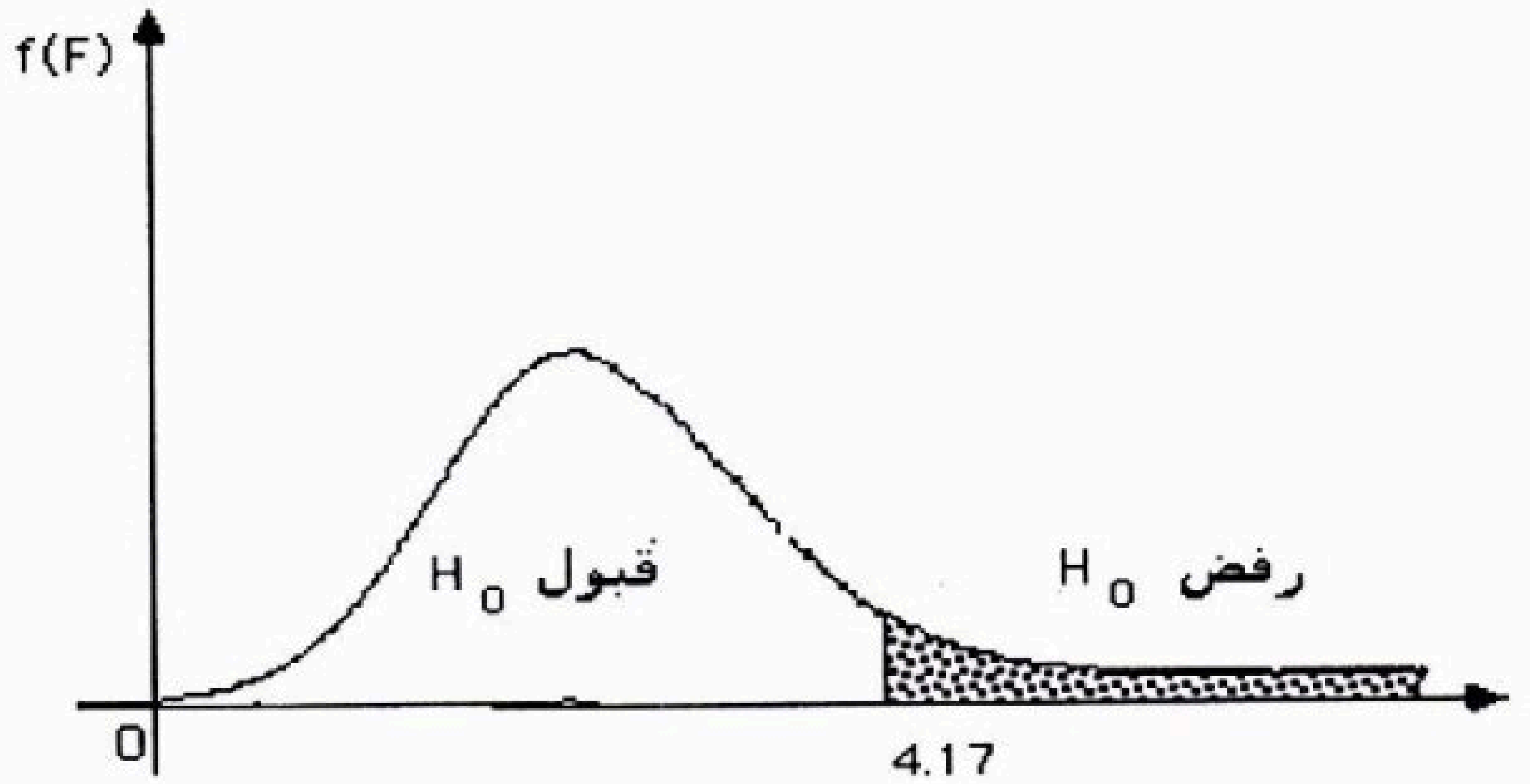
بشرط أن تكون $\upsilon_2 > 2$.

$$V(F) = \frac{2\upsilon_2^2(\upsilon_1 + \upsilon_2 - 2)}{\upsilon_1(\upsilon_2 - 2)(\upsilon_2 - 4)}$$

بشرط أن تكون $\upsilon_2 > 4$.

ونوضح المساحة المظللة في شكل (٤,٢٢) لقيمة الاحتمال α على يمين القيمة $F_{\upsilon_1, \upsilon_2, \alpha}$ في العبارة الاحتمالية (٤,٣٣) كما يلي :

$$P(F_{\upsilon_1, \upsilon_2} \geq F_{\upsilon_1, \upsilon_2, \alpha}) = \alpha \qquad (٤,٣٣)$$



شكل (٤,٢٢)

ولصعوبة عمل جدول التوزيع F لتغير المنحنيات طبقا لتغير درجات الحرية (υ_1, υ_2) فإنه يحسب في جـدول (٤) في آخر الكتاب لقيمتين للاحتمـال α (المساحة المظللة على يمين القيمة $F_{\upsilon_1, \upsilon_2, \alpha}$ شكل (٤ , ٢٢)) أي أن عند $\alpha = 0.05$, $\alpha = 0.01$. وتمثل α قيمة الاحتمال المعطاة بالعلاقة (٤ , ٣٣) على يمين القيـمـة $F_{\upsilon_1, \upsilon_2, \alpha}$ للمتغير $F_{\upsilon_1, \upsilon_2}$ والقيمتين $\alpha = 0.1$, $\alpha = 0.5$ ايضا .

ملاحظة

هناك علاقة تربط قيمتي المتغير F بحيث يكون على يمينه مساحة α أو $1-\alpha$ كما يلي .

$$F_{\upsilon_2, \upsilon_1, 1-\alpha} = \frac{1}{F_{\upsilon_1, \upsilon_2, \alpha}} \qquad (٤ , ٣٤)$$

مثال (٤ , ١٩)

احسب قيم $F_{\upsilon_1, \upsilon_2, \alpha}$ حيث $P(F_{\upsilon_1, \upsilon_2} \geq F_{\upsilon_1, \upsilon_2, \alpha}) = \alpha$ عندما يكون لدينا القيم التالية :

(i) $\upsilon_1 = 9, \upsilon_2 = 12$ ، $\alpha = 0.05$

(ii) $\upsilon_1 = 9, \quad \upsilon_2 = 12, \quad \alpha = 0.95$

(iii) $\upsilon_1 = 15, \ \upsilon_2 = 8$ ، $\alpha = 0.01$

(iv) $\upsilon_1 = 15, \ \upsilon_2 = 8$ ، $\alpha = 0.99$

(v) $P\left(F_{5,7,0.99} \leq F_{5,7} \leq F_{5,7,0.01}\right)$

الحل

من الجدول (٤) آخر الكتاب القيم $\alpha = 0.05, \alpha = 0.01$ نجد أن :

(i) $F_{9, 12, 0.05} = 2.80$

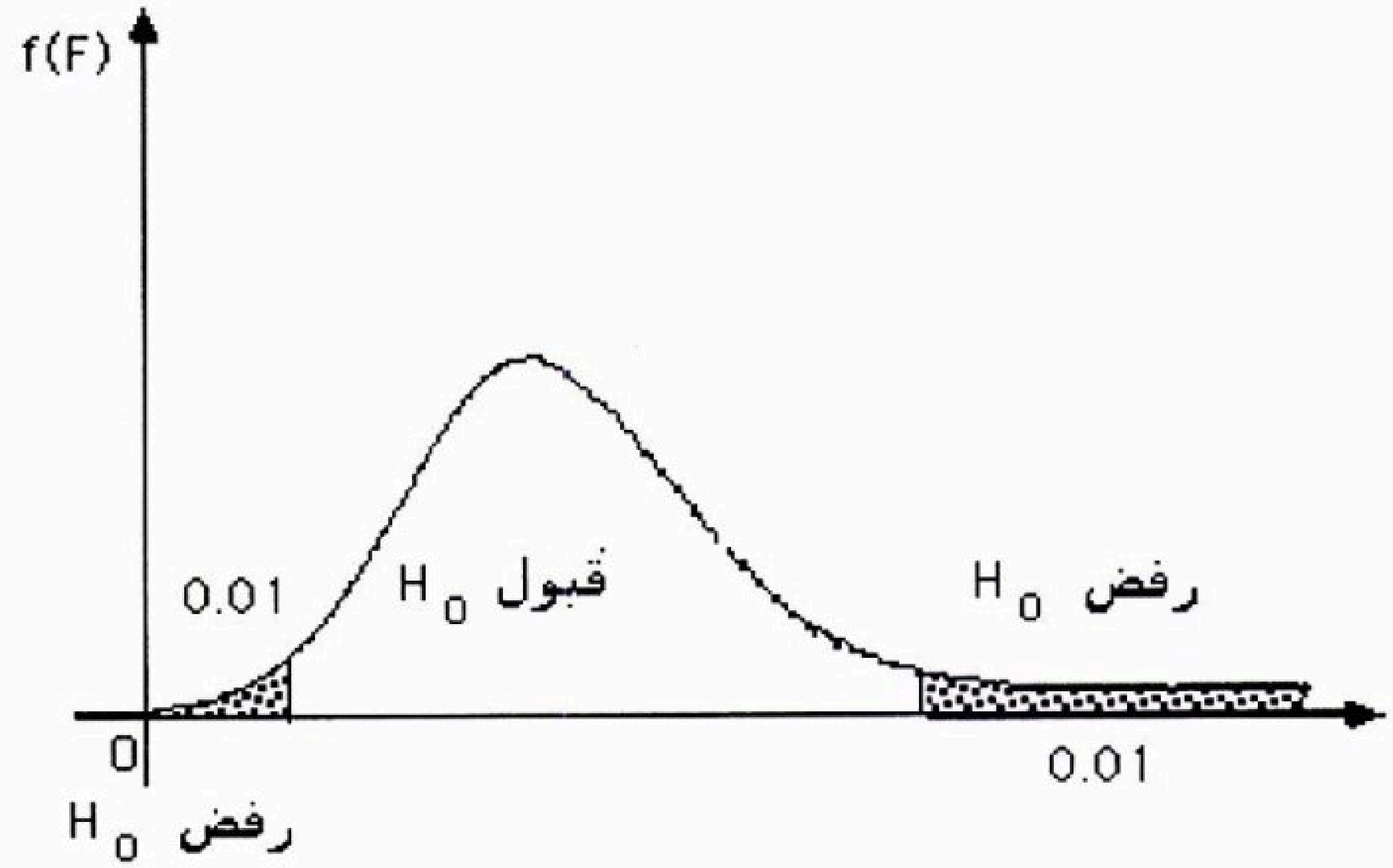
$$F_{9,\,12,\,0.05} = 2.80 \qquad \text{(i)}$$

$$F_{9,\,12,\,0.95} = \frac{1}{F_{12,\,9,\,0.05}} = \frac{1}{3.07} \qquad \text{(ii)}$$

$$F_{15,\,8,\,0.01} = 5.52 \qquad \text{(iii)}$$

$$F_{15,\,8,\,0.99} = \frac{1}{F_{8,\,15,\,0.01}} = \frac{1}{4.00} = 0.25 \qquad \text{(iv)}$$

بحساب قيمة الاحتمال وحدود المتغير في العبارة الاحتمالية في (V) نوضح بالمساحة المظللة في شكل (٢٣ , ٤) التالي الاحتمال خارج المنطقة غير المطلوبة.



شكل (٢٣ , ٤)

$$P(F_{5,\,7,\,0.99} \le F_{5,\,7} \le F_{5,\,7,\,0.01}) = 0.99 - 0.01 = 0.98$$

من جدول (٤) آخر الكتاب نجد أن :

$$F_{5,7,0.99} = \frac{1}{F_{7,5,0.01}} = \frac{1}{10.46} = 0.01$$

$$F_{5,7,0.01} = 8.46$$

ويكون الاحتمال المطلوب هو :

$$P(0.01 \le F_{5,7} \le 8.46) = 0.98$$

(٤,٧) تمارين

١ - أوجد قيمة الثابت k لدالة f(x) حتى تصبح دالة كتلة احتمالية :

(i) $f(x) = kx\ ,\ x = 1,2,3$. (ii) $f(x) = k\left(\frac{3}{x}\right),\ x = 0,1,2,3$.

٢ - إذا كان X متغير عشوائـي، f(x) دالة كتلته الاحتمالية وتعطي فـي الجـدول الآتي :

x	-2	1	0	2	5	8
f(x)	0.13	0.12	0.25	k	0.22	0.03

(i) احسب قيمة الثابت k .

(ii) احسب توقع وتباين المتغير العشوائي X .

(iii) احسب $P(X \le -3), P(X > 9), P(0 \le X \le 5), P(X \ge 2)$

٣ - إذا كان المتغير العشوائى X متصلا ودالة كثافته الاحتمالية f(x) هي

$$f(x) = \begin{cases} kx^2(1-x), & 0 \le x \le 1 \\ 0 & \text{خلاف ذلك} \end{cases}$$

(i) احسب مقدار الثابت k .

(ii) احسب التوقع والتباين للمتغير العشوائى X .

(iii) احسب $P(0.2 \le X \le 0.6)$.

٤ - إذا كان X متغيرا عشوائيا يتبع توزيع ذي الحدين فاثبت أن

$$\mu = n\,p \quad , \quad \sigma^2 = n\,p\,q$$

٥ - إذا كان X متغيرا عشوائيا يتبع توزيع فوق الهندسي فأثبت أن

i) $$\mu = n\,p \;,\; \sigma^2 = n\,p\,q\left(\frac{N-n}{N-1}\right)$$

ii) التوزيع فوق الهندسي $f(x) = \dfrac{\binom{np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}$ يؤول إلي توزيع ذي الحدين $\binom{n}{x} p^{x} q^{n-x}$.

٦ - إذا كان احتمال إصابة قناص لهدف هو 0.7 . فإن صوب نحو هدف 5 مرات متتالية ، وإن عرفنا المتغير العشوائي X بأنه عدد مرات الإصابة .

(i) أوجد دالة الكتلة الاحتمالية للمتغير العشوائي X .

(ii) أوجد توقع وتباين المتغير العشوائي X .

(iii) أوجد احتمال أن يصيب الشخص الهدف مرتين على الأقل .

٧ - شحنة تجارية تحتوي على 60 جهازا كهربائيا من بينها عشرة أجهزة معطلة . اختيرت 5 أجهزة عشوائيا وكان المتغير العشوائي X = عدد الأجهزة المعطلة في العينة المختارة أوجد ما يلي :

(i) دالة كتلته الاحتمالية f(x) للمتغير العشوائي X .

(ii) التوقع والتباين للمتغير العشوائي X .

(iii) احتمال أن يوجد في العينة المختارة جهازا معطلا على الأكثر

٨ - إن كان توزيع بواسون يعطى بالدالة $f(x,\lambda)$ كما يلي:

$$f(x, \lambda) = \frac{\lambda^{x}}{x\,!} e^{-\lambda}, \quad x = 0,1,2,....$$

أوجد الاحتمالات $f(4,0.25),\ f(0,2),\ f(6,2)$

٩ - إن كانت نسبة المصابين بمرض في بلد معين هو 0.004 فما هو احتمال عدم وجود إصابة في حي يسكنه 8000 نسمة .

١٠ - إذاكان المتغير العشوائى Z له توزيع طبيعي قياسي N(0,1) فأوجد باستخدام الجدول (١) آخر الكتاب الاحتمالات التالية :

(i) $\Phi(1.75)\ ;\quad \Phi(-0.65)\quad ;\quad \Phi(0.42)$

(ii) $P(Z < 1.2),\ P(Z > -0.48),\ P(-0.3 \quad Z \quad 0.4)$

١١ - إن كانت درجة الحراة خلال شهر مارس تتبع التوزيع الطبيعي بأحد البلاد بتوقع 22° وانحراف معياري 2.25° . أوجد الاحتمالات

(i) أن تكون درجة الحرارة بين 20.35° , 26.5° .

(ii) أن تكون درجة الحرارة أقل من 18.5° .

(iii) أن تكون درجة الحرارة أكبر من 25° .

١٢ - إن قذفنا قطعة عملة متزنة 90 مرة . باستخدام تقريب ذي الحدين للطبيعي أوجد الاحتمالات التالية :

(i) الحصول على 35 صورة فقط .

(ii) على الأقل الحصول على 35 صورة .

(iii) على الأكثر الحصول على 47 صورة

١٣ - إن كان 60% من السحاب يظهر نموا بأيونات الفضة . أوجد احتمال من بين 80 سحابة عدد 45 سحابة تظهر نموا على الأكثر .

١٤ - أوجد النقاط التالية باستخدام الجداول مع التوضيح بالرسم .

(i) $t_{20,0.05}\,;\,t_{12,0.90}\,;\,t_{20,0.01}\,;\,t_{5,0.05}$

(ii) $\chi^2_{3,0.95};\ \chi^2_{12,0.05};\ \chi^2_{13,0.10};\ \chi^2_{8,0.99}$

(iii) $F_{6,9,0.01}\,;\,F_{8,12,0.05}\,;\,F_{5,7,0.99}$

١٥ - أوجد المقدار k من الجدول مع التوضيح بالرسم

(i) $t_{5,k} = 2.015\ ;\ t_{12,k} = 2.681\ ;\ t_{23,k} = 2.069$

(ii) $\chi^2_{15,\kappa} = 30.578\ ;\ \chi^2_{5,k} = 11.070\ ;\ \chi^2_{15,k} = 5.229$

(iii) $F_{8,9,k} = 3.23\ ;\ F_{9,11,k} = 4.63\ ;\ F_{3,24,k} = 3.72$

الفصل الخامس

المعاينــة وتوزيعاتهــا

Sampling and Its Distributions

● مقدمة ● المجتمعات المنتهية وغير المنتهية ● توزيع المعاينة ● توزيع المعاينة للأوساط ● نظرية النهايــة المركزية ● توزيع المعاينة للنسبــة ● توزيع المعاينــة للتباين ● توزيع المعاينة للأوساط عندما تكـون n صغيرة و σ^2 مجهولة ● توزيعات المعاينة للعينــات المختارة من مجتمعين ● متباينة تشبتشيف ● تمارين

(٥ , ١) مقدمة

Introduction

سبق أن عرفنا المجتمع الإحصائي الذي يتكون من المفردات التي يهـمـنـا دراستها ، وأن هذا المجتمع له بعض الخصائص يمكن حساب بعض المؤشرات له والتي تسمى معالم لهذا المجتمع مثل متوسط المجتمع μ وتباين المجتمع σ^2 ونسبة أفراد المجتمـع P التي تتوافرفيها صفة معينة . . . الخ . سبق أن عرفنا العـيـنـة العشوائية بأنها مجموعة جزئية من مفردات المجتمع محل الدراسة وتختار بطريقة عشوائية بحيث تكون ممثلة لجميع مفردات المجتع محل الدراسة . حساب بعض المؤشرات من هذه العينات العشوائية تسمى إحصاءات وتعتبر تقديرات لمـعـالـم المجتمع الإحصائي محل الدراسة . وهذه الإحصاءات تكون ذات قيم قريبة من المعالم أقل منها قليلا أو أكبر منها قليلا . والفرق بين قيمة الإحصاءه والمعلـمـة يكون راجع إلى الخطأ العشوائي (خطأ الصدفة) . وعلى سبيل المثال، إذا أخذنا

عينتين عشوائيتين من طلاب جامعة الملك سعود حجم كل منهما مائة طالب وكان متوسط الطول للعينة الأولى $\overline{x}_1 = 172$ سم وكان متوسط الطول للعينة الثانية $\overline{x}_2 = 169$ سم . فإن كان معلوم لنا متوسط الطول لمجتمع طلاب الجامعة (معلمة الطول) $\mu = 170$ سم . فإن الزيادة في الإحصاءه الأولى $\overline{x}_1 = 172$ عن معلمة المجتمع $\mu = 170$ سم ناتجه عن خطأ الصدفة ، وكذلك النقص في الإحصاءه الثانية $\overline{x}_2 = 169$ ناتج عن خطأ الصدفة أيضا . ويهمنا في هذا الفصل دراسة توزيع الإحصاءه $\overline{X}$ أو غيرها وتأثير الخطأ العشوائي عليها ، وهذه الدراسة تعرف بما يسمى بتوزيع المعاينة . وسنتناول في هذا الفصل توزيع المعاينة للأوساط، ونظرية النهاية المركزية ، وتوزيع المعاينة للنسبة ، وتوزيع المعاينة للتباين وغيرها ومتباينة تشبتشيف وذلك بالشرح والأمثلة .

(٥,٢) المجتمعات المنتهيةوغيرالمنتهية

Finite and Infinite Populations

في دراستنا لتوزيعات المعاينة يجب أن نفرق بين توزيعات المعاينة المختارة من مجتمع غير منتهي ومجتمع منتهي. وفي حالة العينة المختارة من مجتمع كبير غير منته مثل اختيار عينة من أسماك الخليج العربي، أو عينة من الطيور المهاجرة إلى المملكة في الخريف من الشمال، أو عينة من وحدات إنتاج أحد المصانع، أو سحب عينة بإرجاع من n مصباح من صندوق به عدد محدود من N مصباح عندما يتم السحب بإرجاع فإن المجتمع المنتهي يتحول إلى مجتمع كبير غير منتهي . . . الخ. كما أنه في حالة اختيار عينة من مجتمع محدود ومنتهي مثل اختيار عينة بدون إرجاع من n مصباح من صندوق به N مصباح ، أو اختيار عينه من طلاب قسم الإحصاء بكلية العلوم لهذا العام ، أو اختيار عينة من أعضاء هيئة التدريس بجامعة الملك سعود هذا العام أيضا . . سبب اهتمامنا بالتفرقة بين مجتمع منتهي وغير منتهي راجع إلى اختلاف خصائص توزيع المعاينة للعينات المسحوبة

من المجتمع غير المنتهي والمجتمع المنتهي كما سيوضح خلال دراسة هذا الفصل .

(٥,٣) توزيع المعاينة

Sampling Distribution

إن عدد العينات ذات الحجم n من مجتمع منته ذي حجم N عندما يكون السحب بدون إرجاع هو $\binom{N}{n}$ عينة . أما إذا كان السحب بإرجاع فإن عدد العينات المختلفة هو N^n عينة . فإذا حسبنا من العينات المختلفة إحدى الإحصاءات وليكن الوسط الحسابي $\bar{x}_i$ للعينة i فإننا نحصل على مجتمع جديد مفرداته هي الأوساط الحسابية لهذه العينات المختلفة وأن هذه الأوساط عددها أكبر من عدد مفردات المجتمع الأصلي، ويكون عدد الأوساط (المفردات) للمجتمع الجديد في حالتي السحب كما يلي:

أولا : في حالة السحب بدون إرجاع $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots, \bar{X}_{\binom{N}{n}}$

(مع عدم أخذ الترتيب في الإعتبار)

ثانيا: في حالة السحب بإرجاع $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots, \bar{X}_{N^n}$

وتوزيع المعاينة للإحصاءات سواء كانت أوساط حسابية أو انحراف معياري أو نسبة مفردات عينة لصفة معينة في المجتمع محل الدراسة . . . الخ. يتم دراسة كل منها على حدة فيما يلي :

مثال (٥,١)

مجتمع مكون من 12 مفردة اختيرت عينات مكونة من مفردتين:

ا) أوجد عدد العينات الممكنة ثم اكتب الأوساط الحسابية لهذه العينات وذلك في حالة السحب بدون إرجاع .

ب) أوجد عدد العينات الممكنة ثم اكتب الأوساط الحسابية لهذه العينات وذلك في حالة السحب بإرجاع .

الحل

أ) عدد العينات الممكنة في حالة السحب بدون إرجاع = $\binom{12}{2} = 66$ عينة والأوساط الحسابيه لها

$$\overline{X}_1, \overline{X}_2, \dots, \overline{X}_{66} .$$

ب) عدد العينات الممكنة في حالة السحب بإرجاع = $12^2 = 144$ عينة والأوساط الحسابية لها

$$\overline{X}_1, \overline{X}_2, \dots, \overline{X}_{144} .$$

(١ , ٣ , ٥) العينة العشوائية البسيطة (Simple random sample)

العينة العشوائية البسيطة هي العينة التي حجمها n وتختار من مجتمع منتهي حجمه N بحيث يكون لكل عينة من $\binom{N}{n}$ من العينات المختلفة الاحتمال $1/\binom{N}{n}$ نفسه في الاختيار عندما يكون السحب بدون إرجاع . وأن لكل عينة من N^n من العينات المختلفة الاحتمال $1/N^n$ نفسه في الاختيار عندما يكون السحب بإرجاع .

ويتم سحب عينة عشوائية حجمها n من مجتمع حجمه N بدون إرجاع بالطرق التالية :

ا) تكتب جميع مفردات المجتمع والتي عددها N على بطاقات ثم يتم خلطها جيدا ثم نسحب واحدة بعد الأخرى حتى يتم سحب n بطاقة تحمل أرقام المفردات التي تمثل العينة العشوائية البسيطة . وقد تكون هذه الطريقة غير مناسبة عندما تكون كل من n و N كبيرتين .

ب) باستخدام الأرقام العشوائية يتم اختيار العينة العشوائية البسيطة ويتم توضيح ذلك بمثال كما يلي :

مثال (٥ , ٢)

يعمل في إحدى المؤسسات الصناعية 180 مهندسا و كان المطلوب اختيار 8 مهندسين منهم بطريقة عشوائية .

الحل

يتم استخدام جداول الأرقام العشـوائية ، بأن نرقم المهندسين بأرقام مسلسلة كما يلي :

001, 002, ..., 179, 180

ثم نستخدم الجداول العشوائية رقـم (٥) فـي آخر الكتاب فنحصل على أرقـام للمهندسين على سبيل المثال كما يلي :

042, 005, 179, 164, 037, 153, 041, 018

ملاحظة

عند اختيار هذه الأرقام اخترنا أي ثلاثة أعمدة رأسية من جداول الأرقـام العشوائية لأن المجتمع الأصلي محل الدراسة يتكون من ثلاث خانات وأهملـنـا جميع الأرقام العشوائية التي أكبر مـن 180 وهو حجم المجتمع الأصلي ، كمـا أهملنا الأرقام العشوائية المتكررة .

(٥ , ٤) توزيع المعاينة للأوساط

Sampling Distribution of Means

إذا كان لدينا مجتمع إحصائي محل الدراسة وكان اهتمامنا بدراسـة أحـد المتغيرات له وليكن X = طول أحد الطلاب بجامعة الملك سعود، على سبيل المثال . فإن المتغـيـر الـعـشـوائـي X له مفردات مخـتـلـفـة (أطـوال الـطـلاب) $X_1, X_2, X_3, \dots, X_N$ وهذه المفردات لها متوسط μ (متوسط أطوال طلاب الجامعة) وانحراف معياري σ لهؤلاء الطلاب . فإن اختيار العينات البسيطة الممكنة في حالة السحب بدون إرجاع عددها $\binom{N}{n}$ وبإرجاع يكون عددها N^n . وبحساب

متوسط كل عينة من العينات الممكنة السابقة . فإننا نحصل على مجتمع جديد بمفردات جديدة قيمها تمثل الوسط الحسابي لكل عينة وهذه المفردات قيم للمتغير العشوائى الجديد $\overline{X}$ الذي يأخذ القيم التالية :

أولا : في حالة السحب بدون إرجاع $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_{\binom{N}{n}}$

ثانيا: في حالة السحب بإرجاع $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_{N^n}$

وإن قيم المتغير العشوائي الجديد $\overline{X}$ لها متوسط جديد يرمز له بالرمز $\mu_{\bar{x}}$ وانحراف معياري (خطأ معياري) جديد يرمز له بالرمز $\sigma_{\bar{x}}$ ولدراسة توزيع المعاينة للمتغير $\overline{X}$ وعلاقة متوسطه $\mu_{\bar{x}}$ بمتوسط المجتمع الأصلي μ وعلاقة إنحرافه المعياري $\sigma_{\bar{x}}$ بانحراف المجتمع الأصلي σ باستخدام النظريات الإحصائية والرياضية، فإنه أمكن صياغة نظرية (١ , ٥) التالية :

نظرية (١ , ٥)

لعينات عشوائية بسيطة ذات حجم n مأخوذة من مجتمع له متغير عشوائي X وله متوسط μ وانحراف معياري σ . فإن التوزيع العيني النظري للمتوسط $\overline{X}$ يكون له توقع $E(\overline{X}) = \mu_{\bar{x}}$ وانحراف معياري $\sqrt{V(\overline{X})} = \sigma_{\bar{x}}$ (ويسمى أحيانا خطأ معياري) يعطى بالعلاقات (١ , ٥) ، (٢ , ١٥) و (٢ , ٥ب) كما يلي :

$$\mu_{\bar{x}} = \mu \qquad (٥ , ١)$$

إذا كان المجتمع N صغير (منته) وأن السحب بدون إرجاع

$$\sigma_{\bar{x}} = \left\{ \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right. \qquad (٢ , ١٥)$$

إذا كان المجتمع N كبير (غيرمنته) أوأن السحب بإرجاع من مجتمع صغير

$$\sigma_{\bar{x}} = \left\{ \frac{\sigma}{\sqrt{n}} \right. \qquad (٢ , ٥ب)$$

ملاحظة

أولا: المقدار $\frac{N-n}{N-1}$ يسمى معامل التصحيح ويقترب من الواحد الصحيح كلما كبر حجم المجتمع N مقارنة بحجم العينة n أي أنه عندما تكون النسبة $(\frac{n}{N})$ أقل من 5%.

ثانيا : المعاينة بإرجاع من مجتمع منته تساوي تماما المعاينة بدون إرجاع من مجتمع منته، لأن المعاينة بإرجاع من مجتمع منته يمكن اعتباره من الناحية النظرية مجتمعا غير منته . وعلى سبيل المثال سحبت عينة $n=3$ من الكور من صندوق به $N=10$ من الكور لا ينتهي ما دام السحب من الصندوق بإرجاع ويعتبر مجتمع الكور في الصندوق في هذه الحالة مجتمع غير منته .

مثال (٣ , ٥)

أحد المجتمعات يتكون من العناصر الآتية 1, 3, 4, 5, 7 أوجد التوزيع التكراري العيني للمتغير العشوائي $\overline{X}$ لجميع العينات الممكنة ذات الحجم $n=2$ في كل من الحالتين التاليتين:

أولا: في حالة السحب بدون إرجاع المطلوب التحقق من العلاقتين (١ , ٥) و (٢ , ١٥) في النظرية (١ , ٥) وهما كما يلي:

$$\text{(i)}\ \mu_{\bar{x}}=\mu \quad ; \quad \text{(ii)}\ \sigma^2_{\bar{x}}=\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

ثانيا : في حالة السحب بإرجاع المطلوب التحقق من العلاقتين (١ , ٥) و (٢ , ٥ب) في النظرية (١ , ٥) وهما كما يلي:

$$\text{(i)}\ \mu_{\bar{x}}=\mu \quad ; \quad \text{(ii)}\ \sigma^2_{\bar{x}}=\frac{\sigma^2}{n}$$

الحل

أولا : في حالة السحب بدون إرجاع :

عدد العينات الممكنة هو $\binom{5}{2} = 10$ عينات بيانها كما يلي :

(1,3), (1,4), (1,5), (1,7)
(3,4), (3,5), (3,7)
(4,5), (4,7)
(5,7) .

ومتوسطات هذه العينات هو كما يلي :

2, 2.5, 3, 4
3.5, 4, 5
4.5, 5.5
6

والمتوسطات السابقة للعينات يمكن وضعها في توزيع تكراري كما في جدول (١ , ٥) يمكن منه حساب التوقع $E(\overline{X})$ ويرمز له $(\mu_{\bar{x}})$ والتباين $V(\overline{X})$ ويرمز له $\sigma^2_{\bar{x}}$ كما يلي :

جدول (١ , ٥)

المتوسط $\bar{x}$	التكرار f	التكرار النسبي $f(\bar{x})$	$\bar{x}.f(\bar{x})$	$\bar{x}^2.f(\bar{x})$
2	1	1/10 = 0.1	0.20	0.400
2.5	1	0.1	0.25	0.625
3	1	0.1	0.30	0.900
3.5	1	0.1	0.35	1.225
4	2	0.2	0.80	1.200
4.5	1	0.1	0.45	2.025
5	1	0.1	0.50	2.500
5.5	1	0.1	0.55	3.025
6	1	0.1	0.60	3.600
Σ	10	1	4.00	17.500

من الجدول (١ , ٥) نحسب التوقع $E(\overline{X})$ كما يلي :

$$\mu_{\overline{X}} = E(\overline{X}) = \Sigma\ \overline{x}\ .\ f(\overline{x}) = 4 \tag{٥ , ٣}$$

ومن الجدول (١ , ٥) يحسب التباين $V(\overline{X})$ كما يلي :

$$\sigma^2_{\overline{X}} = V(\overline{X}) = \Sigma \overline{x}^2\ f(\overline{x}) - \mu^2_{\overline{X}}$$

أي أن

$$\sigma^2_{\overline{x}} = 17.500 - (4)^2 = 1.5 \tag{٥ , ٤}$$

ثم نحسب μ, σ للمجتمع الأصلي حيث مفرداته هي 1, 3, 4, 5, 7 كما يلي :

$$\mu = \frac{1+3+4+5+7}{5} = 4 \tag{٥ , ٥}$$

$$\sigma^2 = \frac{(1-4)^2 + (3-4)^2 + (4-4)^2 + (5-4)^2 + (7-4)^2}{5}$$

أي أن

$$\sigma^2 = (9+1+0+1+9)\ /\ 5 = ٤ \tag{٥ , ٦}$$

وبمقارنة العلاقتين (٣ , ٥) و (٥ , ٥) نجد أن $\mu_{\overline{x}} = \mu$

وهذا يحقق العلاقة (١ , ٥) للنظرية (١ , ٥) .

ثم نحسب $\sigma^2_{\overline{x}}$ باستخدام العلاقة (٢ , ٥أ) كما يلي :

$$\sigma^2_{\overline{x}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{4}{2}\left(\frac{5-2}{5-1}\right) = 1.5 \tag{٥ , ٧}$$

وبمقارنة العلاقتين (٤ , ٥) ، (٧ , ٥) نجد أن

$$\sigma^2_{\overline{x}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = 1.5$$

وهذا يحقق العلاقة (٢ , ١٥) للنظرية (١ , ٥). وهو المطلوب أولا .

ثانيا في حالة السحب بإرجاع:

عدد العينات الممكنة هو $5^2 = 25$ عينة بيانها كما يلي :

(1,1), (1,3), (1,4), (1,5), (1,7)
(3,1), (3,3), (3,4), (3,5), (3,7)
(4,1), (4,3), (4,4), (4,5), (4,7)
(5,1), (5,3), (5,4), (5,5), (5,7)
(7,1), (7,3), (7,4), (7,5), (7,7)

ومتوسطات هذه العينات هو كما يلي :

1, 2, 2.5, 3, 4
2, 3, 3.5, 4, 5
2.5, 3.5, 4, 4.5, 5.5
3, 4, 4.5, 5, 6
4, 5, 5.5, 6, 7

والمتوسطات السابقة للعينات يمكن وضعها في توزيع تكراري كما هو موضح في جدول (٢ , ٥) التالي حيث يمكن حساب التوقع $\mu_{\bar{x}}$ والتباين $\sigma^2_{\bar{x}}$ كما يلي:

جدول (٢ , ٥)

المتوسط $\bar{x}$	التكرار f	التكرار النسبي $f(\bar{x})$	$\bar{x}.f(\bar{x})$	$\bar{x}^2.f(\bar{x})$
1	1	$\frac{1}{25} = 0.04$	0.04	0.04
2	2	0.08	0.16	0.32
2.5	2	0.08	0.20	0.50
3	3	0.12	0.36	1.08
3.5	2	0.08	0.28	0.98
4	5	0.20	0.80	3.20
4.5	2	0.08	0.36	1.62
5	3	0.12	0.60	3.00
5.5	2	0.08	0.44	2.42
6	2	0.08	0.48	2.88
7	1	0.04	0.28	1.96
Σ	25		4.00	18.00

من الجدول (٢ , ٥) يحسب التوقع $\mu_{\bar{x}}$ كمايلي :

$$\mu_{\bar{x}} = \sum \bar{x} \,.\, f(\bar{x}) = 4 \qquad (٥ , ٨)$$

وأيضا من الجدول (٢ , ٥) يحسب التباين $\sigma^2_{\bar{x}}$ كما يلي :

$$\sigma^2_{\bar{x}} = \sum \bar{x}^2 \; f(\bar{x}) - \mu^2_{\bar{x}} = 18 - (4)^2 = 2 \qquad (٥ , ٩)$$

بمقارنة العلاقتين (٥ , ٥) ، (٥ , ٨) نجد أن

$$\mu_{\bar{x}} = \mu = 4$$

وهذا يحقق العلاقة (٥ , ١) للنظرية (٥ , ١).

ثم يحسب $\sigma^2_{\bar{x}}$ باستخدام العلاقة (٢ , ٥ب) كما يلي :

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = \frac{4}{2} = 2 \qquad (٥ , ١٠)$$

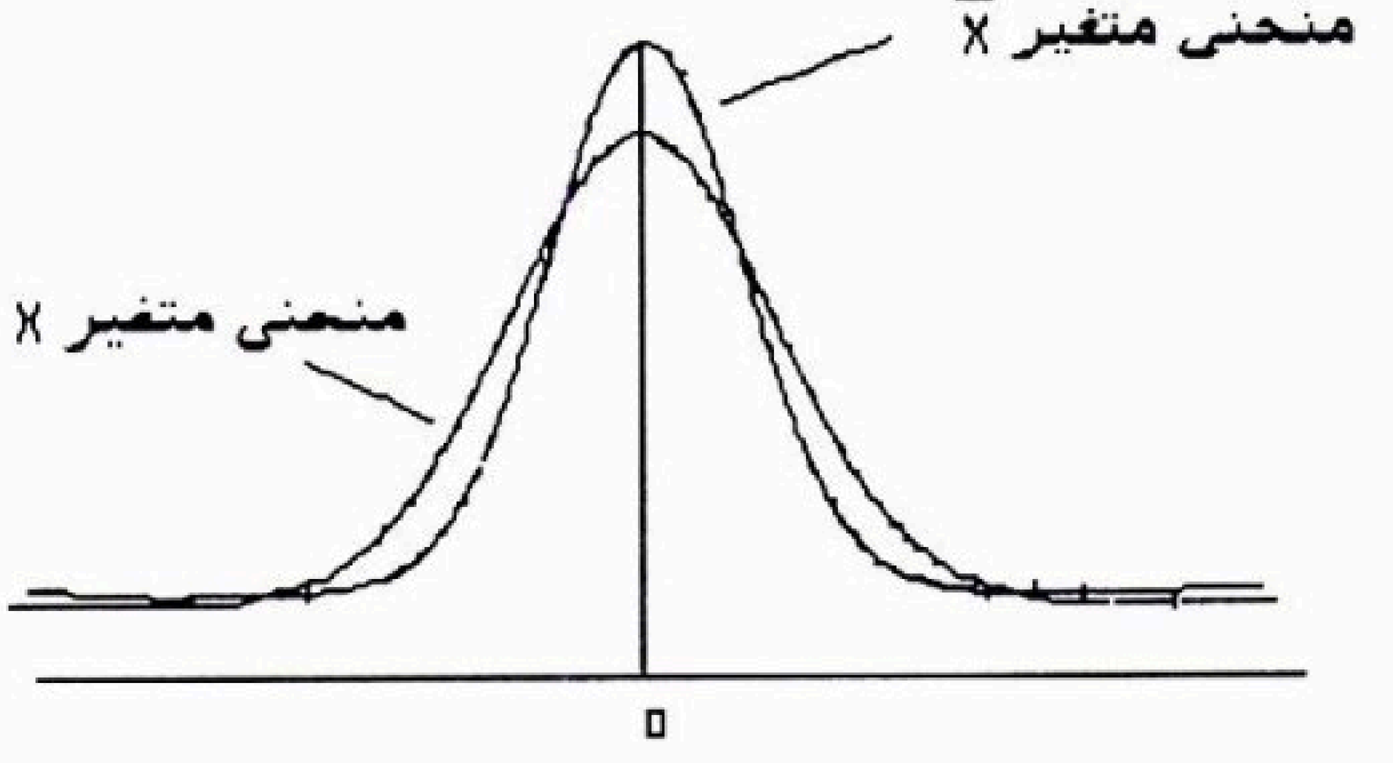
بمقارنة العلاقتين (٥ , ٩) و (٥ , ١٠) نجد أن العلاقة

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = 2$$

وهذا يحقق العلاقة (٢ , ٥ب) للنظرية (٥ , ١). وهو المطلوب ثانيا.

ملاحظة

من نظرية (٥ , ١) نلاحظ أن الانحراف المعياري (الخطأ المعياري) للمتغير العشوائي لمتوسطات العينات $\overline{X}$ مقسوما على $\sqrt{n}$ وهو $(\frac{\sigma}{\sqrt{n}})$ أقل بكثير من الانحراف المعياري للمتغير العشوائي X للمجتمع الأصلي (σ) ، وخاصة عندما يكون حجم العينة n كبيرا نسبيا. ويوضح ذلك في الشكل (٥ , ١) التالي:



شكل (٥,١)

(٥,٥) نظرية النهاية المركزية
(Central Limit Theorem)

إذا اخترنا عينة حجمها n من مجتمع له متوسط μ وانحراف معياري σ وكان المتغير العشوائي $\overline{X}$ لمتوسطات جميع العينات الممكنة المختارة وذات الحجم n ،

فإن المتغير العشوائي الذي له الصيغة $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ له صفتين هما :

(١) يكون له (تقريبا) التوزيع الطبيعي القياسي إذا كانت n 30 مهما كان توزيع المجتمع.

(٢) يكون له (تماما) التوزيع الطبيعي القياسي إذا كان توزيع المجتمع طبيعيا مهما كان حجم العينة n.

ملاحظة

إذا كانت n 30 فإن $\overline{X} \sim N(\mu, \dfrac{\sigma^2}{n})$ وإذا كان المجتمع طبيعيا مهما كانت n فإن $\overline{X} \sim N(\mu, \dfrac{\sigma^2}{n})$.

مثال (٤ , ٥)

فتائل حريرية متوسط نقطة قطعها μ يساوي 15 بانحراف معياري σ يساوي 0.2 كجم . أختيرت عينة من مائة فتيلة وذلك لإيجاد نقطة قطعها . ما هو احتمال متوسط نقطة القطع بحيث تكون $\overline{X}$ داخل الفترة (14.97, 15.02) كجم .

الحل

الاحتمال المطلوب هو :

$$P(14.97 \le \overline{X} \le 15.02)$$

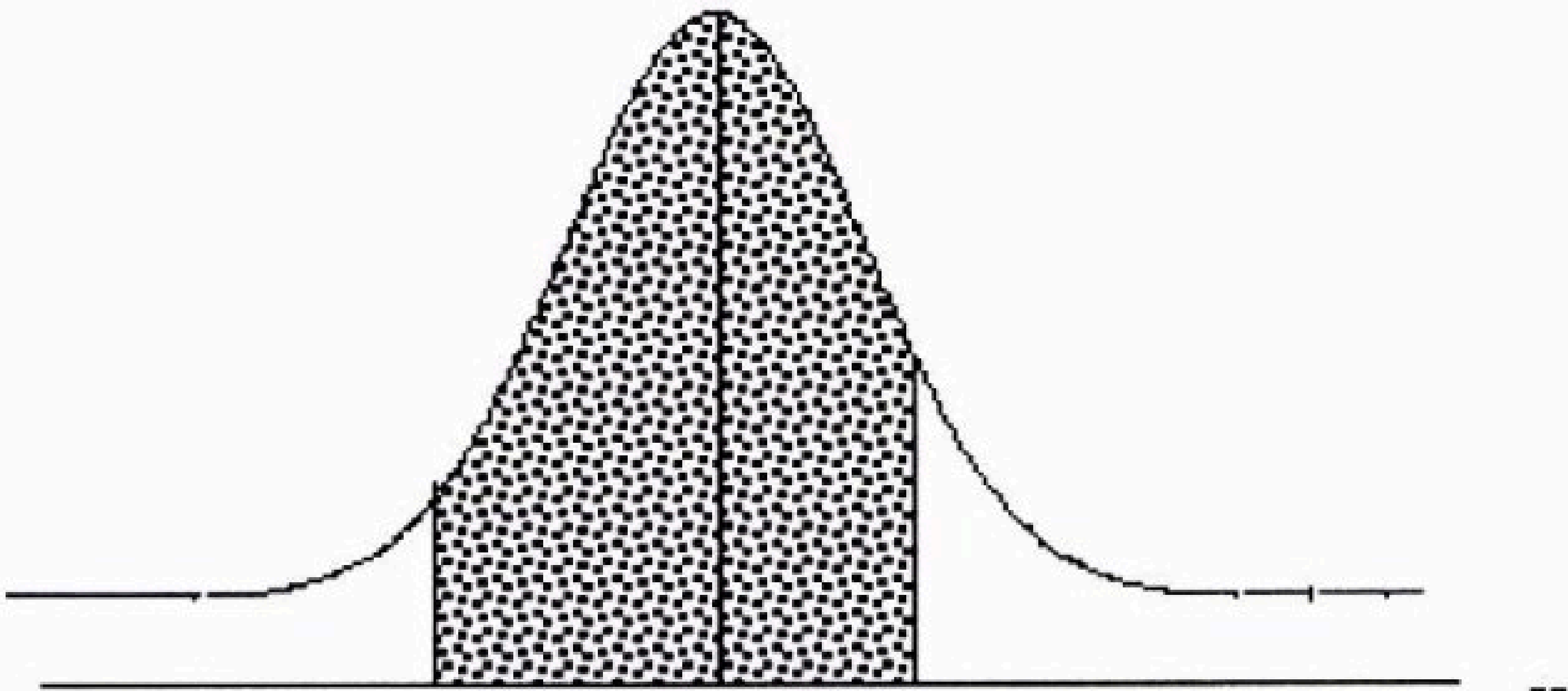
حيث حجم العينة $n = 100$ ، $\mu = 15$ ، $\sigma = 0.2$

وباستـخدام نظرية النهـاية المركزية فإن $\overline{X} \sim N(15, \frac{(0.2)^2}{100})$ يكون الاحتمال المطلوب هو:

$$P(\frac{14.97-15}{0.2/10} \le Z \le \frac{15.02-15}{0.2/10})$$

$$= P(-1.5 \le Z \le 1)$$

ونوضح الاحتمال المطلوب بالجزء المظلل في الشكل (٢ , ٥) كما يلي :



شكل (٢ , ٥)

باستخدام الجدول (١) للتوزيع الطبيعي القياسي آخر الكتاب نجد أن:

$$P(-1.5 \le Z \le 1) = P(0 \le Z \le 1) + \{(0 \le Z \le 1.5)$$
$$= 0.3413 + 0.4332 = 0.7745$$

أي باحتمال 77% تقريبا .

مثال (٥ , ٥)

إذا كانت الحوادث الأسبوعية في أحد شوارع مدينة ما تتبع توزيع بواسون بمتوسط ثلاث حوادث أخذت عينة من 36 أسبوعا . أوجد احتمال أن يكون متوسط عدد الحوادث 3.2 حادثا على الأكثر .

الحل

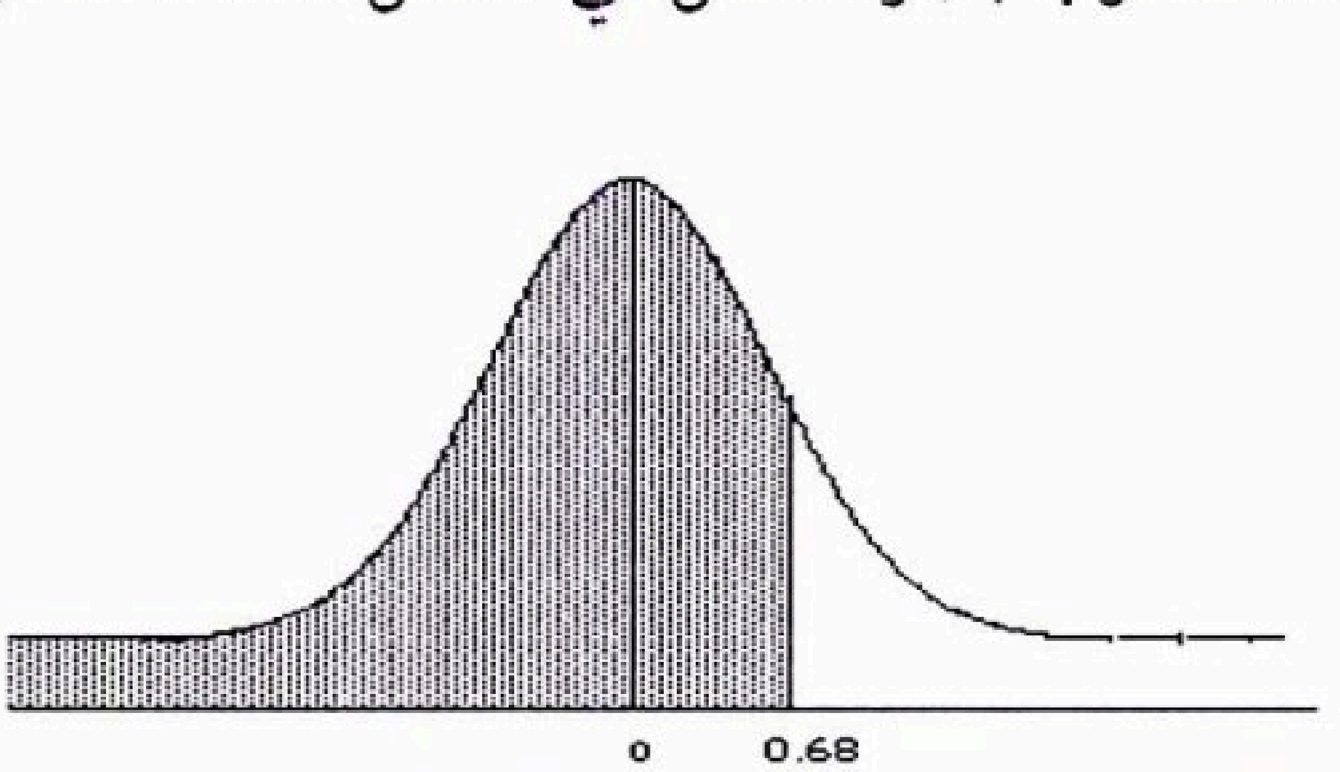
الاحتمال المطلوب $P(\overline{X} \le 3.2)$

وحيث إن المتغير العشوائي X يمثل عدد الحوادث يتبع توزيع بواسون بمتوسط μ وتباين σ^2، فإن $\mu = \sigma^2 = \lambda = 3$.

باستخدام نظرية النهاية المركزية ، فإن الاحتمال المطلوب هو:

$$P(Z \le \frac{3.2 - 3}{\sqrt{3} / 6}) = P(Z \le 0.68)$$

ونوضح الاحتمال المطلوب بالجزء المظلل في الشكل (٣ , ٥) كما يلي:



شكل (٣ , ٥)

وباستخدام الجدول (١) للتوزيع الطبيعي القياسي آخر الكتاب نجد أن

$$P(Z \le 0.68) = 0.5 + P(0 \le Z \le 0.68)$$
$$= 0.5 + 0.2517 = 0.7517$$

أي باحتمال 75% تقريبا .

مثال (٦ , ٥)

إذا كان لدينا إنتاج سلعة بالكجم خلال فترة زمنية (0,3) ساعة رمزنا له بالمتغير العشوائي X الذي دالة كثافته الاحتمالية f(x) هي :

$$f(x) = \begin{cases} \frac{2}{9}x \ , & 0 \le X \le 3 \\ 0 & \text{خلاف ذلك} \end{cases} \qquad (٥ , ١١)$$

اختيرت عينة من 36 وحدة إنتاج لإنتاج هذه السلعة . احسب احتمال متوسط كمية الإنتاج بحيث لا يقل عن 2.1 كجم خلال الفترة (0,3) .

الحل

حيث إن المتغير العشوائي X يمثل كمية الإنتاج خلال الفترة (0,3) فإننا نحسب له المتوسط μ والتباين σ^2 باستخدام دالة الكثافة f(x) من العلاقة (١١ , ٥) كما يلي :

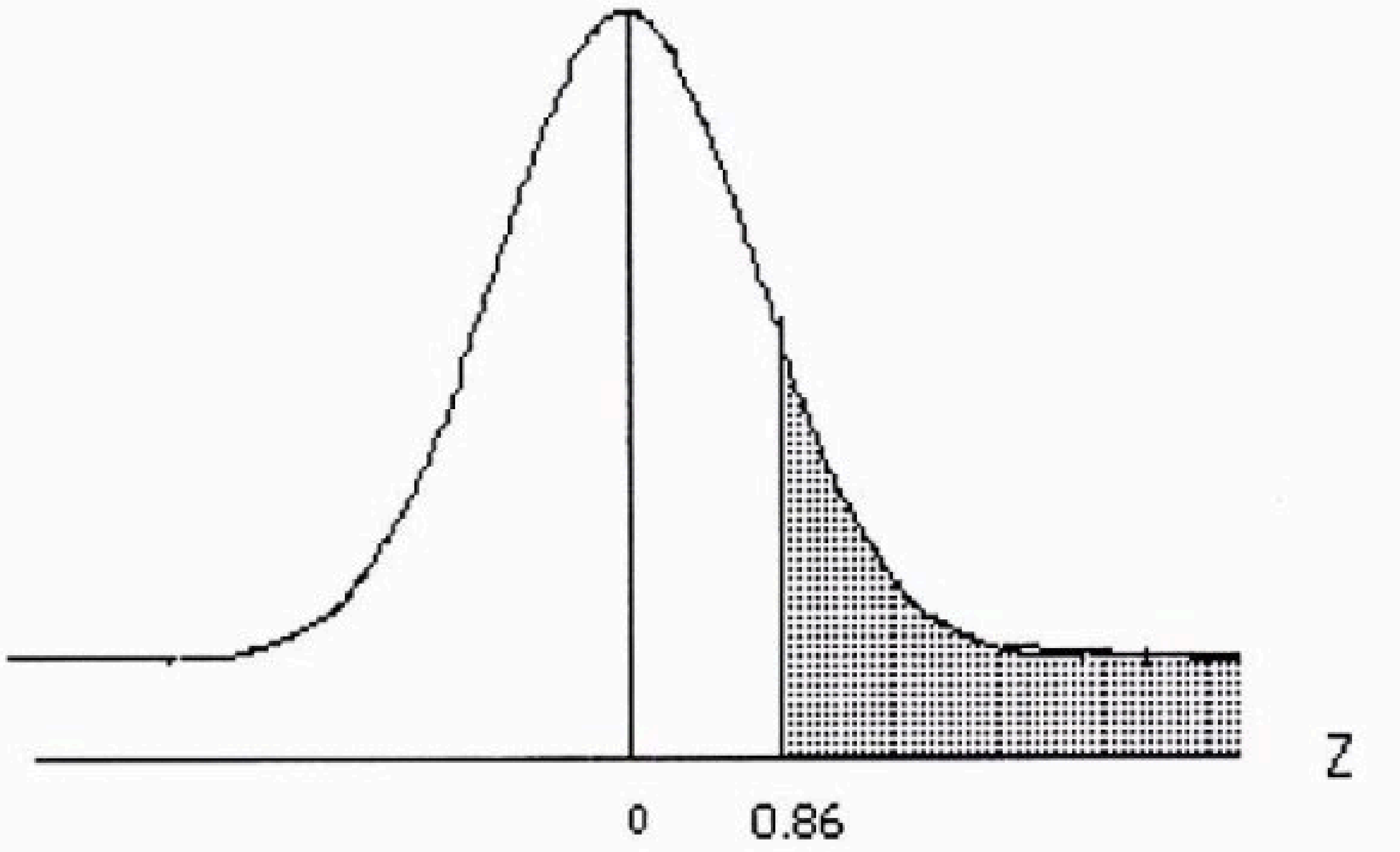
$$\mu = \int_0^{\infty} x.f(x)\, d\,x \ = \int_0^3 x.\frac{2}{9}\, x\, d\,x \ = 2 \quad \text{كجم}$$

$$\sigma^2 = \int_0^{\infty} x^2 f(x)\, d\,x - \mu^2 = \int_0^3 x^2 . \frac{2}{9}\, x\, d\,x - (2)^2 = 0.5$$

والانحراف المعياري $\sigma = \sqrt{0.5} = 0.7$ والعينة $n = 36$. فإن الاحتمال المطلوب هو :

$$P(\overline{X} \geq 2.1) = P(Z \geq \frac{2.1-2}{0.7/6}) = P(Z \geq 0.86)$$

ونوضح الاحتمال المطلوب بالجزء المظلل في الشكل (٤ , ٥) كما يلي:



شكل (٤ , ٥)

من جدول (١) للتوزيع الطبيعي القياسي آخر الكتاب نجد أن :

$$P(Z \geq 0.86) = 0.5 - P(0 \leq Z \leq 0.86)$$
$$= 0.5 - 0.3051 = 0.1949$$

أي باحتمال 19% تقريبا.

مثال (٥ , ٧)

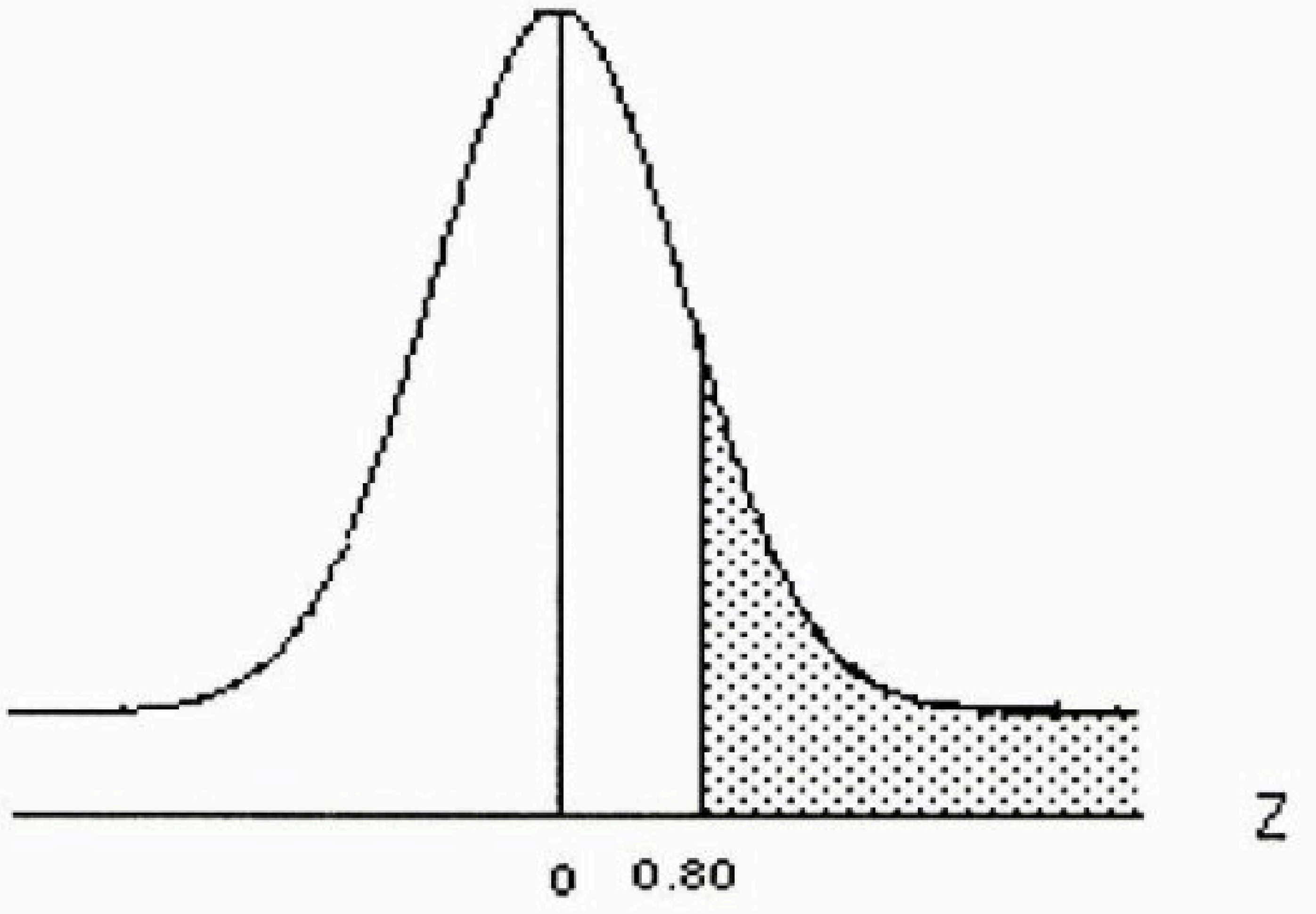
نسبة الذكاء في مجتمع ما يكون لها توزيع طبيعي بمتوسط 100 وانحراف معياري 5 . فإذا اخترنا عينة عشوائية حجمها 16 فردا من هذا المجتمع فما هو احتمال أن يكون متوسط العينة $\overline{x}$ يساوي 101 على الأقل .

الحل

حيث إن المتغير العشوائي X لنسبة الذكاء له توزيع طبيعـي N(100,25) فإن المتغير العشوائي $\overline{X}$ لمتوسط نسبة الذكاء توزيع طبيعي أيضا $N(100\ ,\ \frac{25}{16})$ مهما كان حجم العينة صغيرا $(n = 16)$ ، وأن الاحتمال المطلوب هو

$$P(\overline{X} \geq 101) = P(Z \geq \frac{101 - 100}{5/4}) = P(Z \geq 0.8)$$

ونوضح الاحتمال المطلوب بالجزء المظلل في الشكل (٥,٥) التالي:



شكل (٥,٥)

وباستخدام الجدول (١) للتوزيع الطبيعي آخر الكتاب نجد أن

$$P(Z \geq 0.8) = 0.5 - P(0 \leq Z \leq 0.8)$$
$$= 0.5 - 0.2881 = 0.211$$

أي باحتمال 21% تقريبا .

(٦ , ٥) توزيع المعاينة للنسبة

Sampling Distribution of Proportion

يكون المجتمع الإحصائي أحيانا ذا صفتين فقط . فمثلا عند دراسة ظاهرة التدخين، فإن المجتمع الذي عدد أفـراده N ينقسم قسمين : أشخاص يدخنـون عددهم A فردا ، وآخرون لا يدخنون عـددهـم B فردا . فإننا نعرف متغـيـرا عشـوائيا X كما يلي :

$$X = \begin{cases} 1 & \text{شخص يدخن} \\ 0 & \text{شخص لا يدخن} \end{cases}$$

فإن نسبة الأشخاص الذين يدخنون في المجتمع يرمز لها بالرمـز P وتحسب كما يلي:

$$P = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{A}{N}$$

ونسبة الذين لا يدخنون في المجتمع ويرمز لها بالرمز Q وتحسب كما يلي :

$$Q = 1 - \frac{\sum_{i=1}^{N} x_i}{N} = \frac{B}{N}$$

حيث

$$P + Q = \frac{A}{N} + \frac{B}{N} = \frac{A+B}{N} = 1$$

أي أن

$$Q = 1 - P$$

فإذا أخذنا كل العينات الممكنة ذات الحجـم n من المجتمـع N محل الدراسة ثم

حسبنا لكل عينة النسبة r حيث تمثل نسبة المدخنين في العينة $(r = \frac{a}{n})$ فإنه يصبح لدينا متغيرا عشوائيا جديـدا R له قيــم r_1, r_2, r_3 .... وتوقع μ_R وتباين σ_R^2 وانحراف معياري σ_R حيث إن

$$\mu_R = E(R) = P \ , \ \ V(R) = \sigma_R^2 = \frac{P(1-P)}{n}$$

فإن الإحصاءه المكونة من نسبة العينة وهــي $\frac{r - P}{\sigma_R}$ يكون لها توزيع يقترب مــن N(0,1) وذلك عندما يكون حجم العينـة n كبيرا $(n > 30)$. وذلك حسب نظريـة النهاية المركزية (٥ , ٥) (١) .

مثال (٥ , ٨)

إذا كانت نسبة المعيب في إنتاج إحدى الماكينات هو 20% أخذت عينة من 30 وحدة. أحسب احتمال أن تكون نسبة المعيب 16% على الأكثر .

الحل

نسبة المعيب في المجتمع

$$P = \frac{20}{100} = 0.2 \text{ ،}$$

تباين نسبة المعيب في المجتمع

$$\sigma_R^2 = \frac{P(1-P)}{n} = \frac{0.2(1-0.2)}{30} = 0.0053 \text{ ،}$$

والانحراف المعياري (الخطأ المعياري) هو

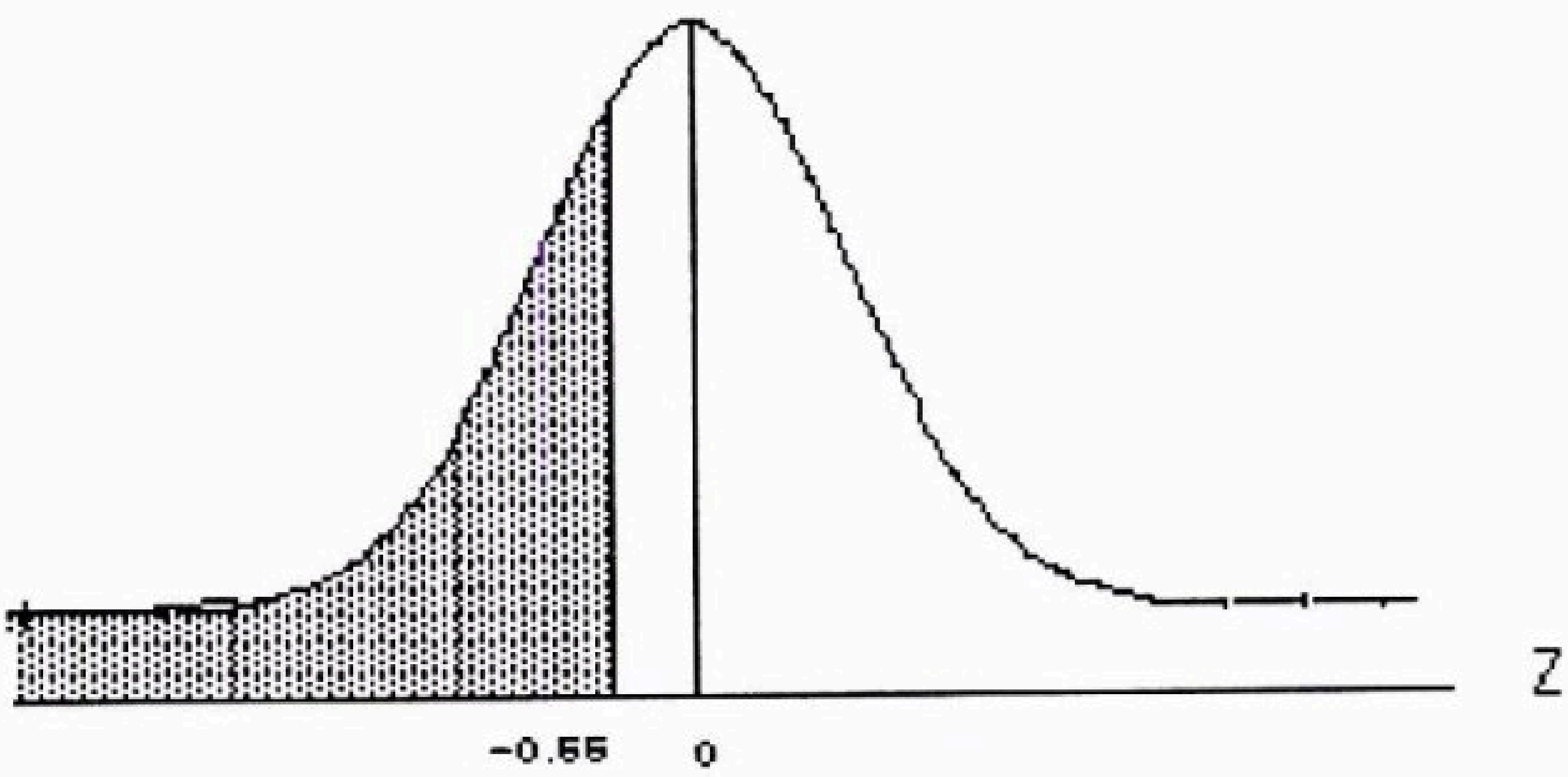
$$\sigma_R = \sqrt{0.0053} = 0.073$$

والإحصاءه $Z = \frac{r-P}{\sigma_R}$ لهاتوزيع يقترب من التوزيع الطبيعي القياسي N(0,1) ومن

ذلك نجد أن الاحتمال المطلوب هو

$$P(r \leq 0.16) = P(Z \leq \frac{0.16 - 0.2}{0.073}) = P(Z \leq -0.55)$$

ونوضح الاحتمال المطلوب بالجزء المظلل في الشكل (٥,٦) كما يلي:



شكل (٥,٦)

ومن جدول التوزيع الطبيعي القياسي (١) آخر الكتاب نجد أن

$$P(Z \leq -0.55) = 0.5 - P(0 \leq Z \leq 0.55)$$
$$= 0.5 - 0.2088 = 0.2912$$

أي باحتمال 29% تقريبا.

(٥,٧) توزيع المعاينة للتباين
Sampling Distribution of Variance

إذا كان لدينا مجتمعا طبيعيا متوسطه μ وتباينه σ^2 ثم حسبنا منه جميع التباينات للعينات العشوائية الممكنة ذات الحجم n ، فإنه يكون لدينا متغيرا عشوائيا جديدا S^2 له قيم هي القيم المختلفة لتباين العينات العشوائية الممكنة

$s_1^2, s_2^2, s_3^2, \ldots\ldots\ldots$ فإن الإحصاءة المكونة من S^2 وهي $\frac{(n-1)\, S^2}{\sigma^2}$ توزيع

χ_ν^2 وذلك بدرجات حريـة $\upsilon = n - 1$ ، وذلك حسب نظرية النهاية المركـزيـة (٥ , ٥) (١).

(٥ , ٨) توزيع المعاينة للأوساط عندما تكون n صغيرة و σ^2 مجهولة

إذا كان لدينا مجتمعا متوسطه μ وتباينه σ^2 مجهولا . وقدرنا تباين σ^2 للمجتمع بتباين العينـة s^2 ذات الحجم n وكانت n صغيرة $(n < 30)$ فإن المتغيـر العشوائي $\overline{X}$ لمتوسطات جميع العينات الممكنة ذات الحجم n والذي يأخذ القيم $\overline{x}_1, \overline{x}_2, \overline{x}_3, \ldots\ldots$. فإنه يمكن صياغة إحصاءة للمتغير العشـوائـي $\overline{X}$ بالصيغة $\frac{\overline{X} - \mu}{s / \sqrt{n}}$ والتي يكون لها توزيـع t بدرجات حرية $\upsilon = n\text{-}1$. وذلك عندما يكون المتغير العشوائي X له توزيع طبيعي $N(\mu,\sigma^2)$ وذلك في حالة التباين σ^2 مجهولا وحجم العينة n صغيرا $(n < 30)$.

(٥ , ٩) توزيعات المعاينة للعينات المختارة من مجتمعين

نفترض أنه لدينا مجتمعين :

أولا : المجتمع الأول متغيرا عشوائيا X_1 ذو متوسط μ_1 وتباين σ_1^2 واخترنا من هذا المجتمع جميع العينات الممكنة ذات الحجـم n_1 . فإنه يتكون لدينا متغيـرا عشوائيا جديدا $\overline{X}_1$ لمتوسطات العينات المختارة يأخذ القيـم $\overline{x}_1, \overline{x}_2, \ldots\ldots$

بتوقع $E(\overline{X}_1) = \mu_1$ وتباين $\sigma^2_{\bar{x}} = V(\overline{X}_1) = \frac{\sigma_1^2}{n_1}$. وعندما يكون حجم كل عينة كبيرا $(n_1 > 30)$ فإنه يمكن حســاب $\sigma^2_{\bar{x}} = V(\overline{X}_1) = \frac{s_1^2}{n_1}$ وذلك باستخدام التباين S_1^2 المحسوب من العينة n_1 وذلك في حالة تباين المجتمع σ_1^2 مجهولا .

ثانيا: المجتمع الثاني المتغير العشوائي X_2 ذو متوسط μ_2 وتباين σ_2^2 . اخترنا من هذا المجتمع جميع العينات الممكنة ذات الحجم n_2 . فإنه يتكون لدينا متغيرا عشوائيا جديــدا $\overline{X}_2$ لمتوسطات العينات المختارة يأخذ قــيــم $\bar{x}_1, \bar{x}_2, \ldots\ldots$ بتوقع $E(\overline{X}_2) = \mu_2$ وتباين $\sigma^2_{\bar{x}_2} = V(\overline{X}_2) = \frac{\sigma_2^2}{n_2}$. عندما يكون حجم كــل عينة كبيرا $(n_2 > 30)$ فإنه يمكن حسابــه $\sigma^2_{\bar{x}_2} = V(\overline{X}_1) = \frac{S_2^2}{n_2}$. حيث إن S_2^2 هو تبــاين العينة ذات الحجم n_2 ، وذلك عندما يكون σ_2^2 للمجتمع مجهولا .

فإنه بواسطة نظرية النهاية المركزية يمكن دراسة المتغيرات العشوائية التالية:

أ) **المتغير العشوائى الجديد $\overline{X}_1 - \overline{X}_2$ من مجتمعين طبيعيين مستقلين** تباينهما σ_2^2 و σ_1^2 معلومين حــيــث $X_2 \sim N(\mu_2, \sigma_2^2)$ و $X_1 \sim N(\mu_1, \sigma_1^2)$.

يكون للمتغير العشوائي الجديد $\overline{X}_1 - \overline{X}_2$ توزيعا طبيعيا أيضا هو :

$$\overline{X}_1 - \overline{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

ب) المتغير العشوائي الجديد $\overline{X}_1 - \overline{X}_2$ من مجتمعين غير طبيعيين مستقلين وحجم كل من n_1 , n_2 كبيرا $(n_1 > 30, n_2 > 30)$. فإن المتغير $\overline{X}_1 - \overline{X}_2$ يتبع التوزيع الطبيعي سواء كانت σ_2^2 و σ_1^2 معلومتين أو مجهولتين أي أن

$$N(\mu_1 - \mu_2 , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$ هو حيث σ_2^2 و σ_1^2 معلومه

أو

$$N(\mu_1 - \mu_2 , \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})$$ هو حيث σ_2^2 و σ_1^2 مجهولة

ج) المتغير العشوائي الجديد $\overline{X}_1 - \overline{X}_2$ من مجتمعين طبيعيين مستقلين وكان $\sigma_1^2 = \sigma_2^2 = \sigma^2$ وكان كل من σ_2^2 و σ_1^2 مجهولة ومقدرة بتباين العينتين S_1^2, S_2^2 حيث كل من n_1 , n_2 صغيرتين فإن المتغير العشوائي الجديد

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ يكون له توزيع t_υ حيث $\upsilon = n_1 + n_2 - 2$ درجات الحرية وتحسب S_p^2 بالمقدار $S_p^2 = \frac{(n_1-1)S_1^2 + (n_1-1)S_2^2}{n_1 + n_2 - 2}$.

د) المتغير العشوائي $\overline{D}$ الذي يحسب من ازواج القيم للعينات المزدوجة من مجتمعين $(x_{11}, x_{21}), (x_{12}, x_{22}) ,..., (x_{1n}, x_{2n})$ طبيعيين غير مستقلين ذات الحجم n في كل من المجتمعين . يحسب المتغير الجديد D حيث $d_i = x_{1i} - x_{2i} \; ; \; i = 1,2,..,n$ ذو متوسط $\mu_d = E(D)$ وتباين هو

$S_d^2 = V(D)$ فإن المتغير العشوائي الجـديـد (أو الإحصاءة) $\frac{\overline{D} - \mu_d}{S_d / \sqrt{n}}$ يكون له

توزيع t_υ حيث $\upsilon = n - 1$ درجات الحرية .

هـ) **المتغير العشوائي الجديد** $R_1 - R_2$ من مجتمعين غير طبيعيين ومستقلين وكل مجتمع ينقسم الى صفتين الصفة الأولى للمجتمع الأول نسبتها في المجتمع P_1 ونسبتهـا r_1 في عينة من هذا المجتمع بحيث يكون حجم هذه العينـة كـبـيـرا $(n_1 > 30)$. نفس الصفة نسبتهـا P_2 للمجتمع الثاني ونسبتهـا r_2 في العينـة المأخوذة منه حيث كان حجم العينة n_2 كبيرا $(n_2 > 30)$. فإن التوقع والتباين للنسبتين r_2 و r_1 للعينتين في كل من المجتمع تحسب كما يلي :

للمجتمع الأول : $E(r_1) = P_1 \,,\; V(r_1) = \frac{P_1(1-P_1)}{n_1}$

للمجتمع الثاني : $E(r_2) = P_2 \,,\; V(r_2) = \frac{P_2(1-P_2)}{n_2}$

فإن الإحصاءة المكونة من الفرق $R_1 - R_2$ تقترب من التوزيـع الطبيعي القياسي كما يلي :

$$Z = \frac{(R_1 - R_2) - (P_1 - P_2)}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}} \sim N(0,1)$$

والإحصاءة المكونة من نسبة تبايني عينتـين S_1^2 , S_2^2 تقترب من توزيـع F كما يلي:

$$\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F_{\upsilon_1, \upsilon_2}$$

بحيث يكون لدينا مجتمعين لكل منهما توزيع طبيعي الأول تباينه σ_1^2 وأخذت منه عينات عشوائية حجم كل منها n_1 ذات تباين S_1^2 . والثاني تباينه σ_2^2 وأخذت منه عينات عشوائية حجم كل منها n_2 ذات تباين S_2^2 .

(٥,١٠) نظرية تشيبتشيف
(Chebyshev's Theorem)

التباين σ^2 أو الانحراف المعياري σ لأي توزيع احتمالي يقيس مقدار التشتت أو انتشار القيم حول المتوسط μ لأي توزيع . فمثلا إذا كانت قيمة الانحراف المعياري σ صغيرة فإن احتمال الحصول على قيمة للمتغير العشوائي قريبة من متوسط التوزيع μ تكون كبيرة جدا وهذه الفكرة تعبر عنها نظرية تشبتشيف والتي تنص على ما يلي:

١) احتمال المتغير العشوائى X يأخذ قيمة x من قيمه الممكنة $X(s)$ تنحرف عن متوسطه μ بمقدار $k\sigma$ على الأكثر يكون على الأقل هو $(1 - \frac{1}{k^2})$ ويعبر عن ذلك بالعبارة الاحتمالية $P(|X-\mu| \le k\sigma) \ge 1 - \frac{1}{k^2}$ حيث الثابت $k > 1$.

أي أن هذه القيمة داخل الفترة $(\mu - k\sigma, \mu + k\sigma)$ وإذا كانت σ^2 و μ مجهولة للمجتمع وقدرت من العينة بالمقدارين s^2 و $\overline{x}$ فإن الفترة السابقة تصبح $(\overline{x} - ks, \overline{x} + ks)$.

(٢) إذا كان المتغير العشوائي $\overline{X}$ لجميع متوسطات العينات لتي حجمها n من المجتمع الأصلي ذو متوسط μ وتباين σ^2 فإن احتمال انحراف القيمة $\overline{X}$ لأي من متوسطات العينات الممكنة تنحرف عن متوسطه μ بمقدار ($k\sigma_{\overline{x}}$ على الأكثر) هو على الأقل $(1 - \frac{1}{k^2})$ ويعبر عنه بالعبارة الاحتمالية:

$$P\left(\left|\overline{X}-\mu\right| \le k\sigma_{\bar{x}}\right) \ge 1-\frac{1}{k^2}$$

حيث الثابت $k > 1$.

أي أن هذه القيمة تقع داخل الفترة $(\mu - k\sigma_{\bar{x}}, \mu + k\sigma_{\bar{x}})$. وأن $\sigma_{\bar{x}}^2$ هو تباين المتغير $\overline{X}$ وتساوي $\sigma_{\bar{x}}^2 = V(\overline{X}) = \frac{\sigma^2}{n}$.

مثال (٩ , ٥)

إذا كان لدينا مجتمعا من طلاب الجامعة له متوسط $\mu = 70$ كجم وانحراف معياري 6 كجم أخذت منه عينة عددها 36 طالبا . اوجد احتمال انحراف متوسطها $\overline{x}$ عن μ حسب العبارة الاحتمالية التالية :

$$P(|\overline{X} - \mu| \le 2)$$

وذلك بطريقتين:

ا) بنظرية تشبيتشيف ب) بنظرية النهاية المركزية .

الحل

ا) بنظرية تشبيتشيف نجد أن :

$$P(|\overline{X}-\mu| \le 2) = P(|\overline{X}-\mu| \le k\frac{\sigma}{\sqrt{n}}) \ge 1 - \frac{1}{k^2}$$

أي يلزمنا حساب الثابت k كما يلي :

$$2 = k\,\frac{\sigma}{\sqrt{n}}$$

$$2 = k\,\frac{6}{\sqrt{36}} \Rightarrow k = 2$$

فإن الاحتمال المطلوب هو :

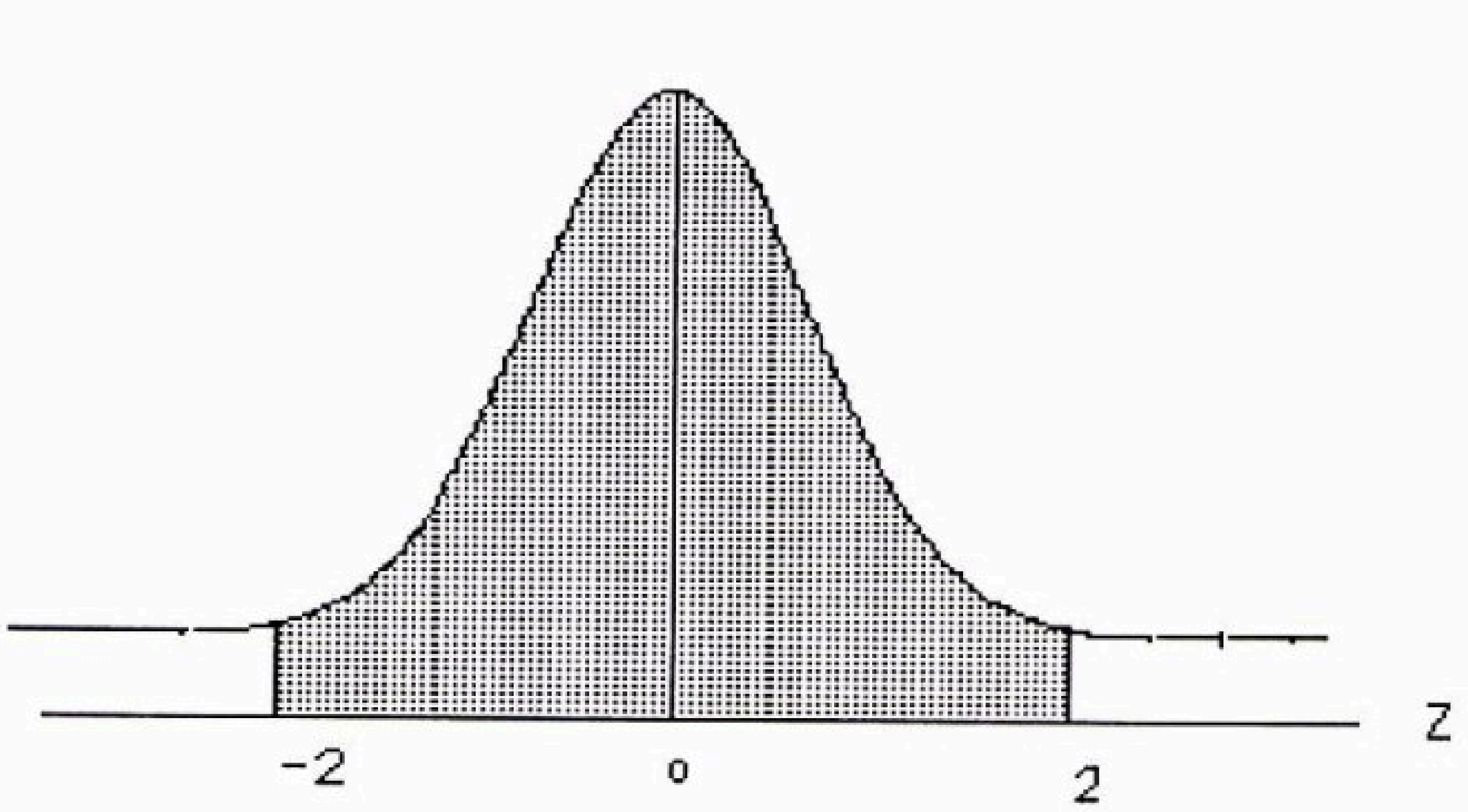
$$P(|\overline{X} - \mu| \leq 2) \geq 1 - \frac{1}{(2)^2} = 0.75$$

ب) بنظرية النهاية المركزية حيث يلزمنا حساب المتغير $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ فإن

الاحتمال المطلوب هو:

$$P(|\overline{X} - \mu| < 2) = P(|\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}| \leq \frac{2}{6/\sqrt{36}}) = P(|Z| \leq 2)$$

ويوضح الاحتمال المطلوب بالجزء المظلل في الشكل (٥,٧) كما يلي:



شكل (٥,٧)

من جدول (١) للتوزيع الطبيعي القياسي آخر الكتاب نجد أن :

$$P(|Z| \leq 2) = 2\,P(0 \leq Z \leq 2)$$

$$= 2(0.4771) = 0.9542$$

(٥ , ١١) تمارين

١- أحد المجتمعات يتكون من العناصر التالـيـة (1,2,3,4,5,6) (مثلا رمي حجر النرد) . أوجد التوزيع العيني للمتغير العشوائي $\overline{X}$ للعينات ذات الأحجام التالية 2,3,4 وتحقق في كل حالة من الحالات السابقة من العلاقتين.

(i) في حالة السحب بدون إرجاع $\mu_{\bar{x}} = \mu,\ \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

(ii) في حالة السحب بإرجاع $\mu_{\bar{x}} = \mu,\ \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

٢ - كم عينة عشوائية بسيطة مختلفة حجمها $n = 2$ يمكن اختيارها بدون إرجاع ثم بإرجاع من مجتمعات محدودة مكونة من

$$N = 20\ ,\ N = 15\ ,\ N = 8\ ,\ N = 5$$

ثم أوجد احتمال اختيار العينات السابقة في حالتي السحب بإرجاع وبدون إرجاع .

٣ - مجتمع يتكون من المفـردات 5,7,9,11 . اكتب له كل العينات العشوائـيـة الممكنة بإرجاع ثم بدون إرجاع المؤلفة من عنصرين من عناصر هذا المجتمع . ثم أوجد توزيع $\overline{X}$ لمتوسط العينة واحسب متوسطه وتباينه $\mu_{\bar{x}}\ ,\ \sigma^2_{\bar{x}}$.

٤ - متوسط عينة عشوائية حجمها $n = 25$ يستخدم كتقدير لمتوسط مجتمع كبير له انحراف معياري $\sigma = 2.5$. ما هو احتمال أن يكون الخطأ أقل مـن 1.2 إذا استخدمنا :

(i) نظرية تشيبتشيف (ii) نظرية النهاية المركزية .

٥ - إذا علمت أن X موزعة توزيعا طبيعيا وسطه 25 وانحرافه المعياري 4 احسب احتمال أن يكون وسط العينة $\bar{x}$ التي حجمها 4 في الحالات التالية :

(i) أكبر من 21 (ii) أكبر من 19.5 (iii) ما بين 21 و 19 .

٦ - إذا كان الانحراف المعياري لأوزان أطفال المرحلة الأولى 6 أرطال . فما هو

احتمال أن يختلف الوزن لمتوسط عينـة $n = 100$ طفلا من هؤلاء الأطفـال بأكثر من 1 رطل عن الوزن لجميع الأطفال .

٧ - إذا كانت أعمار المصابيح الكهربائية المنتجة بواسطة المصانع تتبع توزيعا دالة كثافته الاحتمالية :

$$f(x) = \begin{cases} 12x^2(1-x) \; ; & 0 \le X \le 1 \\ 0 & \text{خلاف ذلك} \end{cases}$$

إذا كان المتغير العشوائي X يمثل عمر المصباح بالسنوات اختيرت عينة حجمها 100 مصباح من إنتاج هذا المصنع احسب احتمال أن يكون متوسط اعمارها 7.2 شهرا على الأكثر باستخدام نظرية النهاية المركزية.

٨ - إذا كان ضغط الدم لمجموعة من الأفراد متوسطـه 90 وانحرافه المعياري 9 واخترنا عينة حجمها 36 فردا من هذه المجموعة . أوجد احتمال أن الوسط الحسابي لضغط الدم في العينة على الأقل 85 وذلك باستخدام:

(i) نظرية تشبتشيف (ii) نظرية النهاية المركزية .

٩ - إذا علم أن نسبة البيض التالف التي تنتجها أحد مراكز إنتاج الدواجن هي 0.03 . اشترى شخص 400 بيضة من إنتاج هذا المركز ، ما هو احتمال أن يجد من بينها 30 بيضة على الأقل تالفة؟ .

١٠ - إذا كانت نسبة المدخنين بين الأفراد الذكور البالغين في إحدى المدن هي 15% اختير خمسون شخصا بطريقة عشوائية ، فما احتمال أن يقل عدد المدخنين بينهم عن عشرة أشخاص .

١١ - إذا كان متوسط أعمار مرضى السكر 55 سنة والانحراف المعياري 20 سنة. اختيرت عينة عشوائية حجمها 36 شخصا مريضا بالسكر . أوجد احتمال:

(i) أن يزيد متوسط العمر في العينة عن 59 سنة .

(ii) أن يقل متوسط العمر في العينة عن 46 سنة .

(iii) أن يتراوح متوسط العمر في العينة بين 57 و 53 سنة .

١٢ - إذا كانت أعمار إحدى السلع بالأسابيع تتبع توزيعا طبيعيا متـوسـطـه 50 أسبوعا وتباينه 25 .

أ - اختيرت إحدى السلع عشوائيا. فما هو احتمال أن يزيد عمرها عن 60 أسبوعا.

ب- اختيرت عينة حجمها 16 سلعة. فما هو احتمال أن يزيد متوسط أعمار السلع في العينة عن 52 أسبوعا. وذلك باستخدام

(١) نظرية تشبيتشيف (٢) نظرية النهاية المركزية.

الفصل السادس

التقدير لمعالم المجتمع الإحصائي

Estimation of Population's Parameters

● مقدمة ● التقدير بنقطة ● التقدير بفترة ● القيمة العظمى للخطأ في التقدير ● تقدير حجم العينة ● تقدير فترة الثقة للمتوسط في العينات الكبيرة ● تقدير فترة الثقة للمتوسط في العينات الصغيرة ● تقدير فترة الثقة للنسبة ● تقدير فترة الثقة للتباين والانحراف المعياري ● تمارين

(١ , ٦) مقدمة

Introduction

تم دراسة بعض التوزيعات الاحتمالية في الفصل الرابع مثل توزيع ذي الحدين حيث له معلمتين n,p، وتوزيع بواسون بمعلمة واحدة λ والتوزيع الطبيعي $N(\mu, \sigma^2)$ بمعلمتين μ, σ^2 وهكذا. يتحدد شكل التوزيع تماما بتحديد قيم هذه المعالم والتي تعتبر ثوابت لأن قيمتها ثابتة لا تتغير. وعادة ما تكون هذه المعالم مجهولة $(\mu, \sigma, P, \ldots)$ للمجتمع. وحيث إن العينة العشوائية تعتبر صورة مصغرة من المجتمع فإننا نلجأ دائما إلى حساب ما يقابل معالم المجتمع من العينة بما يسمى بالإحصاءات (statistics) ومن هذه الإحصاءات الوسط الحسابي $\overline{x}$ للعينة وهو يناظر متوسط المجتمع μ والتباين s^2 للعينة وهو يناظر تباين المجتمع σ^2 والنسبة r للعينة وهي تناظر النسبة P في المجتمع وهكذا.

وتعتبر الإحصاءات $\overline{X}, S^2, R, \cdots$ متغيرات عشوائية لأنها تتغير من عينة إلى

أخرى ذات الحجم نفسه. وسبق لنا دراسة توزيع المعاينة لهذه الإحصاءات ونظرية النهاية المركزية حيث وجدنا الإحصـــاءة $\overline{X}$ للعينة تقترب من التوزيع الطبيـعـي $N(\mu, \frac{\sigma^2}{n})$ في حالة العينات ذات الحجم الكبير $(n > 30)$. والإحصاءة $\frac{\overline{X} - \mu}{S/\sqrt{n}}$ لها توزيع t_υ عندما تكون العينة مأخوذة من مجتمع طبيعي وحجمها صغيرا $(n < 30)$ وتباين المجتمـع σ^2 مجهولا، وهكذا لباقي الإحصاءات. في هذا الفصـل سندرس تقدير المعالم المجـهـولـة $\mu, \sigma, P ,...$ بواسطة الإحصاءات المحسـوبـة $\overline{x}, s, r ,....$ وذلك بالتقدير بنقطة والتقدير بفترة ونوضح ذلك بالشرح والأمثلة كما يلي.

(٦,٢) التقدير بنقطة

Point Estimation

تستخدم قيم الإحصاءات المحسوبة من العينات كتقدير للمعالم المناظرة لها في المجتمع. فمثلا يستخدم متوسط العينة العشوائية $\overline{x}$ كتقدير لمتوسط المجتمـع μ، وكذلك الانحراف المعياري s للعينة كتقدير للانحراف المعياري للمجتـمـع σ وهكذا. . . وتسمى هذه التقديرات التقدير بنقطة لأنها قيمة وحيدة محسوبة من العينة. وعلى سبيل المثال إذا أخذنا عينة عشوائية $(n = 100)$ طالب من جامعة الملك سعود ثم وجدنا متوسط الطول للطلاب في هذه العينة فكان $\overline{x} = 170$ سم، فإنه يمكن القول أن متوسط الطول لمجتمع طلاب جامعة الملك سعود μ يقدر بمتوسط العينة وهو 170 سم. وكذلك إذا أخذنا عينة $(n = 50)$ من إنتاج إحدى الآلات في مصنع وكانت $r = 0.07$ نسبة المعيب للإنتاج في العينة المأخوذة، فإنه يمكن القول إن القيمة التقديرية لنسبة المعيب في إنتاج المصـنـع P هي 7%. إن التقدير بنقطة له خصائص يجب أن تتحقق حتى يصبح تقديرا جيدا مثل عدم التحيز والاتـسـاق وكفؤا وكافيا. . . الخ. ونحن لن ندخل في تفاصيلها في هذا الكتاب لأنها تحتاج

إلى مفاهيم إحصائية متقدمة خارج نطاق هذا الكتاب.

(٦,٣) التقدير بفترة
Interval Estimation

إن التقدير بفترة هو أن نحدد قيمتين تحسبان من العينة العشوائية المأخوذة من المجتمع محل الدراسة. هذه الفترة تحتوي على مجموعة قيم تتضمن فيما بينها قيمة المعلمة المجهولة للمجتمع ويكون حدوث هذه الفترة باحتمال معين $(1-\alpha)$ حيث عادة α تأخذ قيما صغيرة مثل $0.1, 0.05, 0.025, 0.01, ..., 0.005$. واحتمال وقوع المعلمة داخل هذه الفترة يسمى الثقة لأنه يبين ثقتنا بأن المعلمة المجهولة سوف تقع في هذه الفترة. وكلما كان طول الفترة صغيرا زادت دقة التقدير. لذلك سميت بتقدير فترة الثقة. فمثلا إذا كان أكبر فرق للخطأ المطلق بين متوسط المجتمع μ ومتوسط العينة $\overline{x}$ هو المقدار الموجب ε. فإنه يمكن تحديد الاحتمال كما يلي :

$$P(|\mu - \overline{x}| \le \varepsilon) = 1 - \alpha$$

أي أن

$$P(-\varepsilon \le \mu - \overline{x} \le \varepsilon) = 1 - \alpha$$

أي أن

$$P(\overline{x} - \varepsilon \le \mu \le \overline{x} + \varepsilon) = 1 - \alpha \qquad (٦,١)$$

والعلاقة (٦,١) تمثل تقدير لفترة الثقة للمعلمة المجهولة μ، وهذه الفترة هي$(\overline{x} - \varepsilon, \overline{x} + \varepsilon)$. وسندرس في هذا الفصل القيمة العظمى للخطأ في التقدير ويرمز له بالرمز D، وتقدير حجم العينة عندما يكون الخطأ في التقدير معلوما مسبقا، وكذلك تقديرات فترات الثقة للأوساط والنسب والتباين وذلك بالشرح والأمثلة كما يلي :

(٦,٤) القيمة العظمى للخطأ في التقدير

يعرف الخطأ في التقدير بالنسبة لمتوسط المجتمع μ ومتوسط العينة $\overline{x}$ بمقدار الفرق المطلق بينهما ويرمز له بالرمز D ويكتب كما يلي :

$$D = |\overline{x} - \mu| \qquad (٦,٢)$$

ونظرا لاختلاف متوسطات العينات فإن D أيضا يختلف من عينة إلى أخرى معتمدا على متوسط العينة $\overline{x}$. لذلك دعت الحاجة لربط الخطأ في التقدير D المعَرّف بالعلاقة (٦,٢) بعبارة احتمالية وذلك باستخدام نظرية النهاية المركزية كما يلي :

$$P\left(\frac{|\overline{X} - \mu|}{\sigma_{\overline{x}}} \le Z_{\alpha/2}\right) = 1 - \alpha \qquad (٦,٣)$$

أي أن

$$P|\overline{X} - \mu| \le Z_{\alpha/2}\sigma_{\overline{x}}) = 1 - \alpha \qquad (٦,٤)$$

حيث المقدار $D = |\overline{x} - \mu|$ يمثل القيمة العظمى للخطأ في التقدير وأن $\sigma_{\overline{x}}$ تمثل الخطأ المعياري للعينة (الانحراف المعياري لمتوسط العينة) وهو يساوي $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$ للعينات المأخوذة من مجتمعات كبيرة ويساوي $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ للعينات المأخوذة من مجتمعات ذات حجم N وتكون صغيرة بالنسبة لحجم العينة n أي أن النسبة $\frac{n}{N}$ تكون أكبر من 5%.

فمن العلاقة (٦,٤) نجد أن القيمة العظمى للخطأ في التقدير D للعينات

المأخوذة من المجتمعات الكبيرة باحتمال $(1-\alpha)$ هو :

$$D = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \tag{٦,٥}$$

وللمجتمعات الصغيرة ذات الحجم N باحتمال $(1-\alpha)$ هو

$$D = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \tag{٦,٦}$$

مثال (٦,١)

في دراسة عن تلوث الهواء بأوكسيد الكبريت المنبعث من أحد المصانع. سحبت عينة مكونة من 64 يوما، وحسب متوسط العينة فوجد أنه يساوي 16.5 طنا وانحراف معياري 3.5 طنا. احسب وباحتمال 99% مقدار القيمة العظمى الخطأ في التقدير D.

الحل

باستخدام العلاقة (٦,٥) السابقة نجد أن

$$D = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$= 2.58 \frac{3.5}{\sqrt{64}} = 1.13 \quad \text{طنا}$$

(٦,٥) تقدير حجم العينة

عند تقدير القيمة العظمى للخطأ في التقدير (مقدار الدقة المطلوبة) من العلاقتين السابقتين (٦,٥) و (٦,٦) يمكن تقدير حجم العينة بمعلومية القيمة العظمى للخطأ في التقدير كما يلي :

من العلاقة (٦,٥) فإن حجم العينة n_0 يحسب كما يلي :

$$n_0 = \left(\frac{Z_{\alpha/2}\,\sigma}{D}\right)^2 \qquad (٦,٧)$$

ومن العلاقة (٦ , ٦) والعلاقة (٦ , ٧) يمكن حساب حجم العينة n كما يلي:

$$n = \frac{n_0}{1 + n_0/N} \qquad (٦,٨)$$

مثال (٦ , ٢)

ترغب إحدى المؤسسات الصناعية أن تقدر حجم العينة n_0 باحتمال 0.99. وأنها لن تكون مخططة بأكثر من 1.2 في متوسط العينة $\overline{x}$ عن متوسط المجتمع μ. إذا علم أن الانحراف المعياري للمجتمع هو $\sigma = 5$.

الحل

حيث $\sigma = 5$، $D = 1.2$ و $Z = 2.58$ فيكون تقدير حجم العينة n_0 هو

$$n_0 = \left(\frac{Z_{\alpha/2}\,\sigma}{D}\right)^2 = \left(\frac{2.58(5)}{1.2}\right)^2 = 15$$ مفردة تقريبا

(٦ , ٦) تقدير فترة الثقة للمتوسط في العينات الكبيرة

يمكن من العلاقة (٦ , ٤) السابقة حساب تقدير فترة الثقة للمتوسط μ باحتمال $(1-\alpha)$ كما يلي :

$$P\left(\overline{X} - Z_{\alpha/2}\sigma_{\overline{x}} \le \overline{X} - Z_{\alpha/2}\sigma_{\overline{x}}\right) = 1-\alpha \qquad (٦,٩)$$

فإن تقدير حدود فترة الثقة للمجتمعات الكبيرة حيث إن $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ هي كما يلي:

$$\left(\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \;,\; \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) \qquad (٦,١٠)$$

وأن تقدير حدود الثقة للمجتمعات الصغيرة ذات الحجم N حيث إن $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ هي كما يلي:

$$\left(\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \;,\; \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right) \qquad (٦,١١)$$

ملاحظة

يعتبر طول فترة الثقة مساويا ضعف قيمة الحد الأقصى للخطأ D. أي أن طول فترة الثقة = 2D.

مثال (٦,٣)

أوجد من مثال (٦,١) وبدرجة ثقة 99% تقدير فترة الثقة للمتوسط الحقيقي لانبعاث أكسيد الكبريت في الجو من أحد المصانع.

الحل

تقدير فترة الثقة المطلوبة بالعلاقة (٦,١٠) هو

$$\bar{x} - 2.58 \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 2.58 \frac{\sigma}{\sqrt{n}}$$

$$16.5 - 2.58 \frac{3.5}{\sqrt{64}} \le \mu \le 16.5 + 2.58 \frac{3.5}{\sqrt{64}}$$

$$15.37 \le \mu \le 17.63$$

أي أن :

تقدير الحد الأعلى لمتوسط إنبعاث أكسيد الكبريت في الجو=17.63 طنا

تقدير الحد الأدنى لمتوسط إنبعاث أكسيد الكبريت في الجو=15.37 طنا

مثال (٦,٤)

اختيرت عينة عشوائية حجمها 36 عاملا من إحدى المؤسسات الصناعية البالغ عدد العمال بها 500 عامل. فإذا كان متوسط العمر للعامل في العينة هو 35 سنة وانحراف معياري 5 سنوات. فأوجد تقدير فترة ثقة لأعمار العمال بدرجة ثقة 95%.

الحل

نجد في المثال أن حجم المجتمع $N = 500$ عامل، $Z_{\alpha/2} = 1.96$، حجم العينة $n = 36$ ومتوسط عمر العامل $\bar{x} = 35$ سنة. وباستخدام فترة الثقة في العلاقة (٦,١١) نجد أن تقدير فترة الثقة لمتوسط أعمار العمال μ بالمؤسسة هو:

$$\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \le \mu \le \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

وبالتعويض عن قيم $\bar{x}$ ، $Z_{\alpha/2}$ ، σ ، n و N نجد أن:

$$35 - 1.96 \frac{5}{\sqrt{36}} \sqrt{\frac{500-36}{500-1}} \le \mu \le 35 + 1.96 \frac{5}{\sqrt{36}} \sqrt{\frac{500-36}{500-1}}$$

$$35 - 1.6 \le \mu \le 35 + 1.6$$

$$33.4 \le \mu \le 36.6$$

أي أن

تقدير الحد الأعلى لمتوسط أعمار العمال بالمؤسسة = 36.6 سنة تقريبا.

تقدير الحد الأدنى لمتوسط أعمار العمال بالمؤسسة = 33.4 سنة تقريبا.

(٦,٧) تقدير فترة الثقة للمتوسط في العينات الصغيرة

إذا كانت العينة مأخوذة من مجتمع طبيعي $N(\mu, \sigma^2)$ فإن المتغير العشوائي (الإحصاءة) $\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ يتبع التوزيع الطبيعي القياسي $N(0, 1)$ سواء كان حجم العينة n صغيـرا $(n < 30)$ أو كبيرا $(n \geq 30)$ وذلك عندما يكون الانحراف المـعـيـاري σ للمجتمع معلوما. ولكن عندما يكون الانحراف المعياري للمجتمع الطبيـعـي σ مجهولا نستعيض عنه بالانحراف المعياري المقدر من العينة s فإن المتغير العشوائي (الإحصاءة) $\frac{\overline{X} - \mu}{s/\sqrt{n}}$ يقترب من توزيع t_υ وذلك حسب توزيع المعاينة، حيث $\upsilon = n - 1$ تسمى درجات الحرية، وتوزيع t سبق شرحه في الفصل الرابع السابق.

وباستخدام توزيع t يمكن تقدير فترة الثقة للمتوسط μ بدرجة ثقة $(1-\alpha)$ من العبارة الاحتمالية التالية:

$$P\left(\frac{|\overline{X} - \mu|}{s/\sqrt{n}} \leq t_{\upsilon,\alpha/2}\right) = 1 - \alpha \qquad (٦,١٢)$$

أي أن

$$P\left(\overline{X} - t_{\upsilon,\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\upsilon,\alpha/2}\frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

أي أن تقدير فترة الثقة للمتوسط هي :

$$\overline{x} - t_{\nu,\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{\nu,\alpha/2}\frac{s}{\sqrt{n}} \qquad (٦,١٣)$$

وتحسب قيمة $t_{\nu,\alpha/2}$ من جدول (٢) لتوزيع t آخر الكتاب.

مثال (٥ , ٦)

لدراسة اختفاء العلامات الأرضية البيضاء في وسط طريق نتيجة المرور الكثيف أخذت عينة من 6 مناطق مختلفة، ولوحظ اختفاء العلامات البيضاء بعد مرور السيارات مقربة لأقرب ألف سيارة. والبيانات بالآلاف هي :

167, 136, 108, 126, 133, 164

قدر فترة ثقة للمتوسط μ لعدد السيارات اللازمة لاختفاء العلامات البيضاء بدرجة ثقة 0.95.

الحل

يحسب متوسط وتباين العينة $\overline{x}, s^2$ كما يلي :

$\overline{x}$=(167 + 136 + 108 + 126 + 133 + 164 / 6 = 139 ألف سيارة

وأن التباين s^2 يحسب كما يلي :

$$s^2 = \frac{(167-139)^2+(136-139)^2+(108-139)^2+(126-139)^2+(133-139)^2+(164-139)^2}{6-1}$$

$$s^2 = \frac{(28)^2 + (-3)^2 + (-31)^2 + (-13)^2 + (-6)^2 + (25)^2}{5}$$

$$s^2 = \frac{2584}{5} = 516.8$$

وأن الانحراف المعياري s هو :

$$s = \sqrt{516.8} = 22.733$$

وباستخدام العلاقة (١٣ , ٦) يمكن تقدير فترة الثقة للمتوسط μ وايجاد قيمة $t_{5,\,0.025}$ من جدول (٢) لتوزيع t آخر الكتاب نجد أن:

$$139 - t_{5,\,0.025}\ \frac{22.733}{\sqrt{6}} \le \mu \le 139 + t_{5,\,0.025}\ \frac{22.733}{\sqrt{6}}$$

$$139 - (2.571)\left(\frac{22.733}{\sqrt{6}}\right) \le \mu \le 139 + (2.571)\left(\frac{22.733}{\sqrt{6}}\right)$$

$$139 - 23.861 \le \mu \le 139 + 23.861$$

$$115.139 \le \mu \le 162.861$$

أي أن

تقدير الحد الأعلى لمتوسط عدد السيارات هو 162861 سيارة تقريبا.

تقدير الحد الأدنى لمتوسط عدد السيارات هو 115139 سيارة تقريبا.

مثال (٦ , ٦)

في تجربة على 9 من رواد الفضاء في مجال يحاكي مجال انعدام الوزن وجد أن متوسط زيادة ضربات القلب لهم هو 26.25 دقة بانحراف معياري 3.35 دقة في الدقيقة. احسب أقصى خطأ في التقدير للمتوسط لزيادة الضربات μ بدرجة معنوية $\alpha = 0.01$. ثم أوجد حدود الثقة للمتوسط μ.

الحل

نحسب أقصى خطأ في تقدير المتوسط μ من العلاقة (١٢ , ٦) هو

$$D = |X - \mu| = t_{\upsilon,\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$$

$$= t_{8,\,0.005}\ \left(\frac{3.35}{\sqrt{9}}\right)$$

$$= (3.355)\ (1.12) = 3.7464 \quad \text{دقة}$$

وأن تقدير فترة الثقة للمتوسط μ لمستوى معنوية $\alpha = 0.01$

$$\overline{X} - D \le \mu \le \overline{X} + D$$

$$26.25 - 3.7464 \le \mu \le 26 + 3.7464$$

$$22.5036 \le \mu \le 29.9964$$

تقدير الحد الأعلى لمتوسط زيادة ضربات القلب في الدقيقة29.9976 دقة أي 30دقة تقريبا

تقدير الحد الأدنى لمتوسط زيادة ضربات القلب في الدقيقة 22.5036 دقةأي 22.5 دقة تقريبا .

(٨ , ٦) تقدير فترة الثقة للنسبة

سبق دراسة المتغير العشوائى R من العينات ذات الحجم n من مجتمع كبير له معلمة P تمثل نسبة الظاهرة في المحتمع، على سبيل المثال نسبة ظاهرة التدخين في المجتمع. فإنه باستخدام نظرية النهاية المركزية عندما يكون حجم العينـات n كبيرا $(n \ge 30)$. فإن الإحصاءة $\frac{r - P}{\sigma_r}$ يكون لها توزيع معاينة يقترب من التوزيع الطبيعي القياسي N(0,1) حيث σ_r الخطأ المعياري للنسبة في العينة ويحسب مـن المجتمعات الكبيرة بالقيمة $\sigma_r = \sqrt{\frac{r\,q}{n}}$ وفي المجتمعـات المحدودة ذات الحجم N حيث تكون النسبة $\frac{n}{N}$ أكبر من %5 وقيمته $\sigma_r = \sqrt{\frac{r\,q}{n}}\ \sqrt{\frac{N-n}{N-1}}$، حيـث إن

$q = 1 - r$. يحسب تقدير فترة الثـقـة P للمجتمع وبدرجة ثـقـة $(1-\alpha)$ بالعبارة الاحتمالية التالية:

$$P\left(\frac{|r-P|}{\sigma_r} \le Z_{\alpha/2}\right) = 1 - \alpha \qquad (٦,١٤)$$

وتحسب القيمة العظمى للخطأ في التقدير D كما يلي :

$$D = |r - P| \le Z_{\alpha/2}\ \sigma_r$$

وأن تقدير فترة الثقة للنسبة P للمجتمع عند مستوى معنوية α كما يلي :

$$r - D \le P \le r + D \qquad (٦,١٥)$$

أي أن

$$r - Z_{\alpha/2}\sqrt{\frac{r\,q}{n}} \le P \le r + Z_{\alpha/2}\sqrt{\frac{r\,q}{n}} \qquad (٦,١٦)$$

وذلك في المجتمعات الكبيرة.

وأن

$$r - Z_{\alpha/2}\sqrt{\frac{r\,q}{n}}\sqrt{\frac{N-n}{N-1}} \le P \le r + Z_{\alpha/2}\sqrt{\frac{r\,q}{n}}\sqrt{\frac{N-n}{N-1}} \qquad (٦,١٧)$$

للمجتمعات المحدودة، وأن نسبة حجم العينة n إلى حجم المجتمع N لا تقل عن 5% أي أن$(\frac{n}{N} \ge 5\%)$.

مثال (٦,٧)

في دراسة لعينة مكونة من 150 شخصا في أحد المصايف لمعرفة رأيهم عند

تفضيلهم هذا المصيف عن غيره أجاب عدد قدره 108 منهم بأن سبب التفضيل هو دفء هذا المصيف. احسب مقدار القيمة العظمى للخطأ في التقدير $D = |r - P|$ عند درجة ثقة 0.99 ثم أوجد تقدير حدود الثقة للمعلمة P لنسبة المصطافين في المجتمع.

الحل

يحسب من العلاقة (٦,١٤) قيمة أقصى خطأ في التقدير D كما يلي:

$$D = |r - P| = Z_{\alpha/2}\ \sigma_r$$

$$\sigma_r = \sqrt{\frac{r\,q}{n}}\ ,\ Z_{0.005} = 2.58\ ,\ q = 1\text{-}0.72 = 0.28\ ,\ r = \frac{108}{150} = 0.72$$

أي أن

$$D = 2.58\sqrt{\frac{0.72(0.28)}{150}} = 0.095$$

وأن تقدير فترة الثقة لنسبة التفضيل P للمجتمع هي:

$$r - D \le P \le r + D$$

$$0.72 - 0.095 \le P \le 0.72 + 0.095$$

$$0.625 \le P \le 0.815$$

أي أن

تقدير الحد الأعلى لنسبة التفضيل هو 82% من الأشخاص المصطافين تقريبا

تقدير الحد الأدنى لنسبة التفضيل هو 63% من الأشخاص المصطافين تقريبا

ملاحظة

يمكن ايجاد تقدير فترات الثقة للفرق بين متوسطين (μ_1, μ_2) وذلك في حالة العينات الكبيرة $(n_1, n_2 \ge 30)$ والعينات الصغيرة $(n_1, n_2 < 30)$ عندما

يكون الانحراف المعياري لكل منهما مجهولا ويقدر من العينتين بالمقدارين s_2^2, s_1^2 حيث كانت المجتمعات كبيرة كما يلي :

ا) العينات الكبيرة

$$(\bar{x}_1-\bar{x}_2)-Z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \le \mu_1-\mu_2 \le (\bar{x}_1-\bar{x}_2)+Z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \quad (٦,١٨)$$

ب) العينات الصغيرة :

$$(\bar{x}_1-\bar{x}_2)-t_{\alpha/2,\upsilon}\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1-\mu_2 \le (\bar{x}_1-\bar{x}_2)+t_{\alpha/2,\upsilon}\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \quad (٦,١٩)$$

حيث إن $\upsilon = n_1 + n_2 - 2$

وإن

$$s_p^2 = \frac{(n_1-1)\,s_1^2+(n_2-1)\,s_2^2}{n_1+n_2-2}$$

ج) كذلك يمكن تقدير فترة الثقة للفرق بين النسبتين $P_1 - P_2$ كما يلي :

$$(r_1-r_2)-Z_{\alpha/2}\sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}} \le P_1-P_2 \le (r_1-r_2)+Z_{\alpha/2}\sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}} \quad (٦,٢٠)$$

(٦,٩) تقدير فترة الثقة للتباين والانحراف المعياري

سبق أن وجدنا تقديرا لتباين المجتمـع σ^2 بقيمة تباين العينة s^2 وهو ما يسمى بالتقدير بنقطة ومنه يمكن حساب تقدير الانحراف المعياري للمجتـمـع σ بقيمة الانحراف المعياري للعينة s. وحيث التباين للعينة s^2 يعتبر متغيرا عشوائيا

(إحصاءة) له توزيع معاينة، حيث أمكن صياغة متغير عشوائي منه يعطي بالقيمة $\frac{(n-1)S^2}{\sigma^2}$ له توزيع مربع كاي χ^2_υ بدرجات حرية $\upsilon = n - 1$ وباستخدام المتغير العشوائي الجديد $\frac{(n-1)S^2}{\sigma^2}$ ومعرفة توزيعه χ^2_υ أمكن بناء تقديرا لفترة ثقة لتباين المجتمع σ^2 عند درجة ثقة $(1-\alpha)$ من العبارة الاحتمالية التالية :

$$P(\chi^2_{\upsilon,1-\alpha/2} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{\upsilon,\alpha/2}) = 1-\alpha \qquad (٦,٢١)$$

ويحسب من العلاقة (٦-٢١) قيمة تقدير فترة الثقة للتباين σ^2 كما يلي :

$$\frac{(n-1)s^2}{\chi^2_{\upsilon,\alpha/2}} \le \sigma \le \frac{(n-1)s^2}{\chi^2_{\upsilon,(1-\alpha/2)}} \qquad (٦,٢٢)$$

ويحسب تقدير فترة الثقة للانحراف المعـيـاري σ بأخذ الجذر للعلاقـة (٦,٢٢) السابقة فنحصل على :

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\upsilon,\alpha/2}}} \le \sigma \le \sqrt{\frac{(n-1)s^2}{\chi^2_{\upsilon,(1-\alpha/2)}}} \qquad (٦,٢٣)$$

مثال (٦,٨)

احسب تقدير فترة الثقة للتباين σ^2 ثم الانحراف المعياري σ لعدد السيارات اللازمة لاختفاء العلامات البيضاء في العينة المأخوذة لمرور السيارات في 6 مناطق مختلفة في مثال (٦,٥) السابق.

الحل

من مثال (٦،٥) السابق كان لدينا المعلومات التالية :

تباين العينة $s^2 = 516.8$

عدد المناطق $n = 6$

درجات الحرية $\upsilon = n - 1 = 5$

مستوى المعنوية $\alpha = 0.05$

فإن تقدير فترة الثقة لتباين المجتمع σ^2 لعدد السيارات للمناطق المختلفة يعطى من العلاقة (٦،٢٢) كما يلي :

$$\frac{(6-1)516.8}{\chi^2_{5,\,0.025}} \le \sigma^2 \le \frac{(6-1)516.8}{\chi^2_{5,\,0.975}}$$

من جدول (٣) لتوزيع مربع كاي آخر الكتاب نجد قيمة كل من

$$\chi^2_{5,\,0.025} = 12.832 \;,\; \chi^2_{5,\,0.975} = 0.831$$

أي أن

$$\frac{5(516.8)}{12.832} \le \sigma^2 \le \frac{5(516.8)}{0.831}$$

$$201.4 \le \sigma^2 \le 3110$$

وأن تقدير فترة الثقة للانحراف المعياري σ لعدد السيارات للمناطق المختلفة يعطى من العلاقة (٦،٢٣) كما يلي :

$$\sqrt{201.4} \le \sigma \le \sqrt{3110}$$

$$14.192 \le \sigma \le 55.767$$

أي أن

تقدير الحد الأعلى للانحراف المعياري لعدد السيارات للمناطق المختلفة هو 55767 سيارة.

تقدير الحد الأدنى للانحراف المعياري لعدد السيارات للمناطق المختلفة هو 14192 سيارة.

(٦,١٠) تمارين

١ - مصنع لإنتاج المصابيح الكهربائية، اختيرت من إنتاجه عينة عشوائية من 50 مصباحا فإذا كان متوسط عمر المصباح في العينة هو 1300 ساعة وانحرافه المعياري 200 ساعة. أوجد تقديرا لفترة الثقة لمتوسط أعمار إنتاج المصنع كله عند مستوى معنوية $\alpha = 0.05$ و $\alpha = 0.01$.

٢ - إذا كان متوسط طول عينة مكونة من 300 ورقة من نبات الغار هو 142 ملم. وكان معلوما أن الانحراف المعياري لطول الأوراق هو 16 ملم. فأوجد تقدير حدود الثقة لمتوسط طول أوراق نبات الغار عند درجة ثقة 0.95.

٣ - لدراسة النمو لنوع خاص من الزهور أخذت عينة من 90 زهرة وجد أن متوسط النمو خلال العام هو 34.5 سم والانحراف المعياري هو 3.7 سم. أوجد حدود الثقة للمتوسط الحقيقي للنمو عند مستوى معنوية $\alpha = 0.01$.

٤ - لدراسة تشمل 300 عائلة وجد أن متوسط ما أنفق على الطعام خلال عام هو 4200 ريالا بانحراف معياري 450 ريال. أوجد تقديرا لفترة الثقة للمتوسط عند درجة ثقة 0.95.

٥ - ما هو حجم العينة المطلوب لنتوصل إلى حقيقة أن متوسط العينة يختلف عن متوسط المجتمع بأقل من 0.45 وذلك باحتمال 0.95 إذا كان الانحراف

المعياري يساوي 2.30 .

٦ - في اختبار للزمن الذي يستغرقه تجميع ماكينة معينة وجد أن الزمن الذي يستغرقه 6 ماكينات هو على التوالي :

14, 15, 13, 7, 13, 10 دقيقة

أوجد تقدير فترة الثقة للمتوسط عند مستوى معنوية 0.01.

٧ - أخذت العينة 2.4 ,0.7 ,1.2 ,1.1 ,1.8 من توزيع طبيعي $N(\mu, 1)$. أوجد فترة الثقة للمتوسط μ.

٨ - اختيرت عينة حجمها 50 طالبا من طلاب كلية الطب البالغ عددهم 800 طالب فإذا كان متوسط عمر الطالب 22 سنة وانحراف معياري 3 سنوات. أوجد فترة ثقة لمتوسط عمر الطالب في الكلية، بدرجة ثقة 99%.

٩ - اختيرت عينة عشوائية حجمها 500 حذاء من إنتاج مصنع لإنتاج الأحذية فوجد أن من بينها 80 حذاء معيبا. أوجد تقديرا لفترة الثقة لنسبة الأحذية المعيبة في الإنتاج كله عند درجة ثقة 0.95.

١٠ - إذا كان ضغط الدم لمجموعة من الأفراد يتبع توزيعا طبيعيا، اختيرت عينة عشوائية حجمها عشرة أفراد فكان ضغط الدم لمفرداتها هي:

100	95	98	112	122
108	96	102	91	111

استخدم هذه البيانات لايجاد تقدير فترة الثقة لمتوسط ضغط الدم للمجموعة كلها عند مستوى معنوية $\alpha = 0.05$.

١١ - اختيرت عينة مكونة من 500 طالب من طلاب الجامعة وجد من بينهم 50 طالبا يستخدمون اليد اليسرى في الكتابة. أوجد تقديرا لفترة الثقة لنسبة الطلاب الذين يستخدمون يدهم اليسرى في الكتابة عند درجة الثقة 99%.

١٢ - إذا كانت أوزان طلاب إحدى المدارس تتبع توزيعا طبيعيا. اختيرت عينة حجمها 16 طالبا من هذه المدرسة فكانت أوزانهم بالكيلوجرام كما يلي :

41	46	51	43	44	46	52	51.2
46.1	44	50.2	45	46	50	40.5	48

(i) أوجد تقدير فترة الثقة لمتوسط أوزان طلاب هذه المدرسة كلها باستخدام درجة ثقة 0.95.

(ii) أوجد تقدير فترة الثقة لتباين أوزان طلاب هذه المدرسة كلها باستخدام درجة ثقة 0.95.

(iii) أوجد تقدير فترة الثقة للانحراف المعياري لطلاب المدرسة عند مستوى معنوية $\alpha = 0.01$.

الفصل السابع

اختبارات الفروض الإحصائية

Testing Statistical Hypotheses

● مقدمة ● صياغة الفرض الإحصائي ● تحديد مستوى المعنوية ● إجراء الاختبار الإحصائي ● اتخاذ القرار الإحصائي ● اختبار الفروض للمتوسط في العينات الكبيرة ● اختبار الفروض للمتوسط في العينات الصغيرة من مجتمع طبيعي ● اختبار الفروض للنسبة في المجتمع ● اختبار الفروض لتباين المجتمع ● اختبار الفروض للفرق بين متوسطي مجتمعين ● اختبار الفروض للفرق بين نسبتي مجتمعين ● تمارين

(١ , ٧) مقدمة

Introduction

يحاول الإنسان في كثير من الأحيان اتخاذ قرار بشأن ظاهرة معينة في مجتمع ما، وذلك بناء على بيانات عينة عشوائية اختيرت من المجتمع نفسه، ونوضح ذلك بالأمثلة التالية:

(ا) يدّعي باحث اجتماعي أن نسبة التدخين بين الأشخاص البالغين في مجتمع هو أقل من 15%، ويريد التأكد من صحة هذا الإدعاء. فإذا سأل جميع الأشخاص في المجتمع للتعرف على أنهم يدخنوا أم لا فإن ذلك يكون أمر شاق بالنسبة للباحث. لذا يلجأ الباحث لأخذ عينة من المجتمع ثم يحسب نسبة التدخين في العينة فإذا وجدها تمثل 12% في العينة. فهل إدعاء الباحث الاجتماعي صحيحا أم نسبة التدخين في المجتمع أقل 15%. إذا استطعنا بطريقة ما وبشيء

من الثقة أن نوضح أن هذا الفرق لا يمكن أن يكون راجعا لمجرد الصدفة فإن هذا يعني أن نسبة التدخين في المجتمع تقل بكثيرعن 15% .

(ب) يدّعي باحث زراعي بأن أحد الأبقار في مجموعة من المزارع بأحد المناطق، أن متوسط إنتاج اللبن اليومي للبقرة هو أكثر من 25 كجم. ويريد التأكد من صحة الإدعاء فإنه يلزم حساب متوسط إنتاج اللبن اليومي لجميع الأبقار بالمزارع المختلفة وهذا يلزمه جهد شاق. ولذا يلجأ بأخذ عينة عشوائية من الأبقار يحسب منها متوسط إنتاج اللبن اليومي للبقرة فإذا وجده 27 كجم. فهل ادعاء الباحث صحيحا أم متوسط إنتاج اللبن اليومي للبقرة أكبر من 25 كجم. فإذا استطاع الباحث بطريقة ما وبشيء من الثقة أن يوضح أن هذا الفرق ليس ناتج عن الصدفة، فإن هذا يعني أن الفرق في المتوسط ناتج من أن متوسط إنتاج اللبن اليومي للبقرة أكبر من 25 كجم.

(جـ) يدّعي باحث مالي بمؤسسة أن متوسط الدخل الشهري للعامل من مصنع (I) لا يختلف عن متوسط الدخل الشهري للعامل من مصنع (II) والتابعين لهذه المؤسسة. وللتأكد من صحة إدعاءه فإنه يلزمه حساب متوسط الدخل الشهري لجميع العمال في كل من المصنعين (I) ,(II) والمقارنة بينهما وعملية حصر جميع العمال صعبة ومجهدة بالنسبة للباحث. ولذلك فإنه يضطر إلى أختيار عينة عشوائية من بين العمال في كل من المصنعين ويحسب متوسط الدخل الشهري لكل مصنع، فمثلا إذا كان متوسط الدخل الشهري للعامل في المصنع (I) من العينة هو 2250 ريال ومتوسط الدخل الشهري للعامل في المصنع (II) من العينة هو 2200 ريال. فهل هذا يعني أن متوسط الدخل الشهري للعمال بالمصنع (I) يختلف عن العمال بالمصنع (II) أم الفرق ناتج عن الصدفة في العينة. والسؤال هنا هل هذا الاختلاف ناتج عن الصدفة أم عن اختلاف حقيقي بين دخول العمال في المصنعين. . .

من الأمثلة السابقة في ا، ب، ج يمكننا ايجاد معنى للمفاهيم التالية - الفرض الإحصائي -- الاختبار للفرض الإحصائي - درجة الثقة - مستوى المعنوية.

فمثلا ادعاء الباحث الاجتماعي في (ا) أن نسبة التدخين في المجتمع تساوي 15% فأقل هو فرض إحصائي. وكذلك ادعاء الباحث الزراعي أن متوسط إنتاج اللبن اليومي للأبقار يســاوي 25 كجم فأكثر هو فرض إحصائي وادعاء الباحــث المالي بأن متوسط الدخل الشهري بالمصنعين I و II للعمال غيرمختلف هو أيضـا فرضا إحصائيا. والأسلوب الذي بواسطته نستطيع الحكم على صحــة الـفـرض الإحصائي يسمى بالاختبـار الإحصائي للفرض، أي أن الاختبار الإحصائي هــو مجموعة قواعد يمكن بواسطتها قبول أو رفض الفرض. ومقدار ثقتنا في القــرار المتخذ بالرفض أو القبول تسمى درجة الثقة، كما أن مقدار عدم الثقة في القــرار المتخذ بالقبول أو الرفض يسمى مستوى المعنوية.

(٢ , ٧) صياغة الفرض الإحصائي
Statistical Hypothesis

تنقسم الفروض الإحصائية إلى قسمين:

(i) فروض حول معالم لمجتمع (Parametric hypothesis)

(ii) فروض عن شكل دالة التوزيع (Non parametric hypothesis)

ونكتفي في هذا الفصل بدراسة الفروض حول معالم المجتمع فقط. حيث إن الفرض الإحصائي هو إدعاء أو تخمين معين حول معلمة من معالم المجتمــع يكون المطلوب اختبار صحة هذا الإدعاء. وعادة يصاغ الفرض الإحصائي بشكل عدم وجود اختلاف أو عدم وجود علاقة ويسمى بفرض العــدم (null hypothesis) ويرمز له بالرمز H_0. ومن مثال (ج) السابق فإن إدعاء الباحث المالي يصاغ بفرض العدم كما يلي:

H_0: نفترض عدم اختلاف متوسط الدخل الشهري للعامل في المصنعين.

وإلى جانب فرض العدم H_0 يوجـد فـرض بديـل (alternative hypothesis) ويرمز له بالرمز H_1 وهذا الفرض يجب أن يكون صحيحا في حالة عدم صحــة H_0. نجري الاختبار وتكون نتيجته إما رفض H_0 أو قبوله. فإذا كان القرار قبول H_0 في

مثال (جـ) هذا يعني لا يوجد اختلاف في متوسط الدخل بين عمال مصنع (I) ومصنع (II) وأن الاختلاف ناتج عن الصدفة في العينتين وليس حقيقي.

وعدم رفض H_0 لايعني بالضرورة أن الفرض H_0 صحيحا ولكن لا توجد مبررات تقودنا إلى عدم صحتة. كما أن رفض H_0 يعني أن الفرض H_0 خطأ. وللتأكد من صحة وخطأ H_0 يلزمنا دراسة جميع مفردات هذا المجتمع محل الدراسة وهذه عملية شاقة وعالية التكاليف. لذلك نلجأ إلى دراسة عينة عشوائية بدلا من دراسة المجتمع. فإذا كانت نتائج العينة تتفق مع الفرض H_0 فإننا لا نستطيع رفضه وإذا كانت تختلف مع الفرض H_0 فإننانرفضه.

وفيما يلي نوضح فرض العدم H_0 والفرض البديل H_1 في الأمثلة ا، ب، جـ السابقة كما يلي:

أولا: نفرض أن نسبة التدخين في المجتمع هي P. فإنه حسب المثال (ا) يكون فرض العدم H_0 والفرض البديل H_1 كما يلي:

$$H_0 : P = 0.15$$

$$H_1 : P < 0.15$$

ثانيا: نفرض أن متوسط إنتاج اللبن اليومي للبقرة هو μ. فإنه حسب مثال (ب) يكون فرض العدم H_0 والفرض البديل H_1 كما يلي:

$$H_0 : \mu = 25 \text{ كجم}$$

$$H_1 : \mu > 25 \text{ كجم}$$

ثالثا: نفرض أن متوسط الدخل الشهري للعمال في مصنع (I) هو μ_1 وأن متوسط الدخل الشهري للعمال في مصنع (II) هو μ_2 وحسب مثال (جـ)، فإن فرض العدم H_0 والفرض البديل H_1 كما يلي:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

(٧,٣) تحديد مستوى المعنوية
Level of Significance

عند اختبار فرض العدم H_0 ضد الفرض البديل H_1 نجد أنفسنا أمام أربع حالات نلخصها في جدول (٧,١) كما يلي:

جدول (٧,١)

H_0 خطأ	H_0 صحيح	فرض عدم القرار
خطأ من النوع الثاني ويرمز له بالرمز β	قرار سليم	قبول H_0
قرار سليم	خطأ من النوع الأول ويرمز له بالرمز α	رفض H_0

واحتمال الوقوع في خطأ من النوع الأول (أي أن نرفض H_0 وهو صحيح) يسمى مستوى المعنوية (level of significance) ويسمى احيانا بحجم منطقة الرفض (size of rejection region) ويرمز له بالرمر α أي أن α = احتمال رفض H_0 وهو صحيح.

واحتمال الوقوع في الخطأ من النوع الثاني (أي أن قبول H_0 وهو خطأ) ويرمز له بالرمز β أي أن β = احتمال قبول H_0 وهو خطأ، والاختبار الجيد الذي يمكنه تصغير كل من الخطأين α و β معا في وقت واحد وهذا صعب، ولذا يلجأ الإحصائيون إلى تثبيت α (مايسمى بمستوى المعنوي) وعادة قيم α الشائعة في التطبيق هي $\alpha = 0.01$ و $\alpha = 0.05$. فمثلا إذا كانت $\alpha = 0.01$ فهذا يعني احتمال الوقوع في الخطأ من النوع الأول (رفض H_0 وهو صحيح) هو 0.01.

(٧,٤) إجراء الاختبار الإحصائي
Testing of Statistical Hypothesis

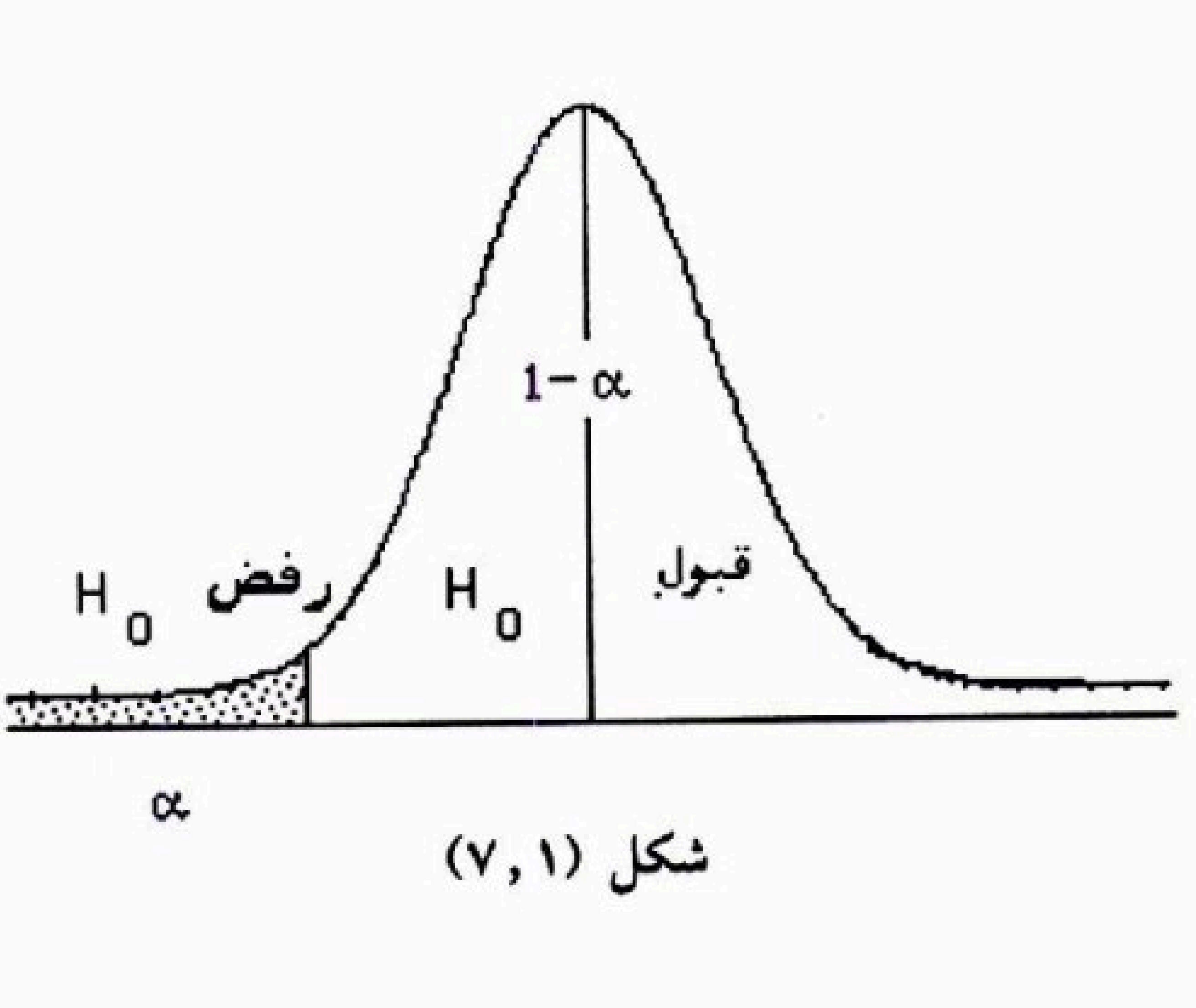
لاختبار صحة فرض العدم H_0 يجب علينا تكوين إحصاءة وهي دالة من

مشاهدات العينة العشوائية وعادة ما يكون توزيع هذه الإحصاءة معلوما. وتقسم القيم الممكنة للإحصاءة (حيث تعتبر متغيرا عشوائيا) إلى قسمين: **القسم الأول** وهو ما يسمى بمنطقة قبول H_0 وهي المنطقة التي يكون احتمال حدوث قيم الإحصاءة فيها كبيرا $(1-\alpha)$ حيث يكون فيها فرض العدم H_0 صحيحا. **والقسم الثاني** يسمى منطقة رفض H_0 وهي التي يكون فيها احتمال حدوث قيم الإحصاءة صغيرا أو نادرا (α) وحيث إن فرض العدم H_0 صحيحا. فيما يلي نوضح مناطق الرفض والقبول لفرض العدم H_0 في الأمثلة ا، ب، ج السابقة كما يلي:

أولا: في مثال أ نسبة التدخين 15% فأقل. فإن H_0 و H_1 يصاغا كما يلي:

$$H_0 : P = 0.15$$
$$H_1 : P < 0.15$$

فإن منطقة رفض H_0 وقيمتها α تمثل بالمساحة المظللة شكل (٧,١) على يسار المنحنى كما يلي:



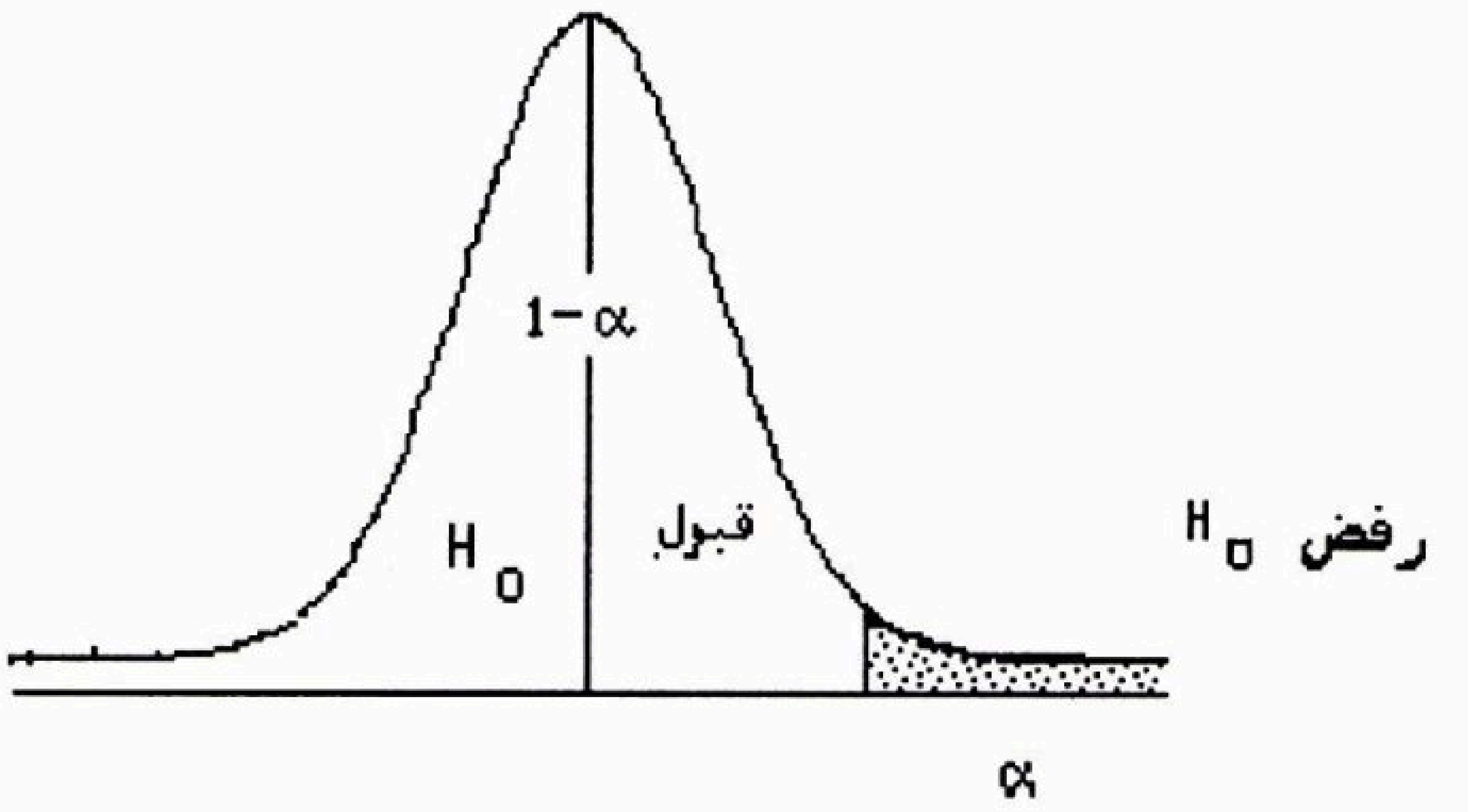
شكل (٧,١)

ثانيا: في مثال (ب) متوسط إنتاج اللبن اليومي للبقرة هو 25 كجم فأكثر فإن H_0 و H_1 كما يلي:

$H_0 : \mu = 25$ كجم

$H_1 : \mu > 25$ كجم

فإن منطقة رفض H_0 وقيمتها α الممثلة بالمساحة المظللة شكل (٧,٢) على يمين المنحنى كما يلي:



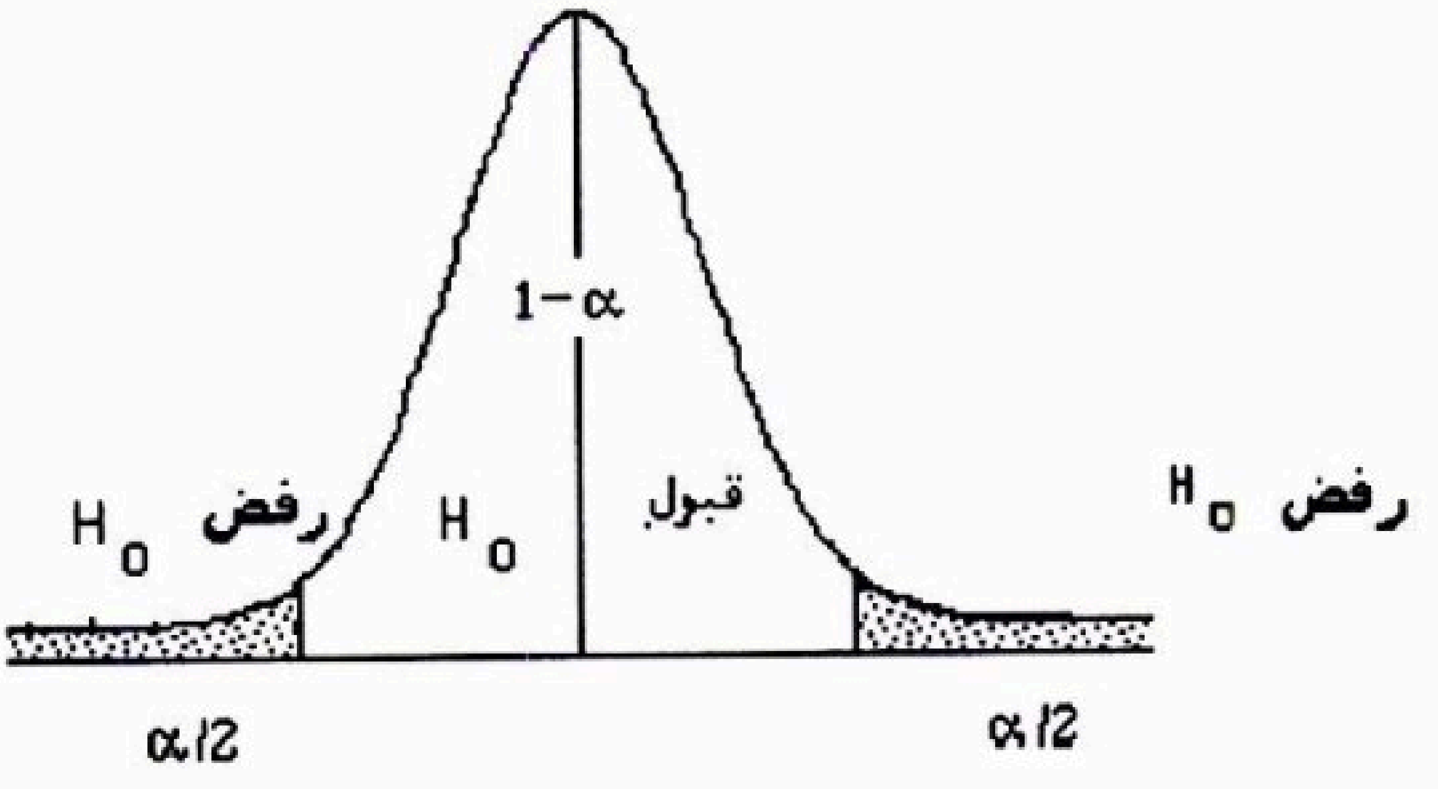
شكل (٧,٢)

ثالثا: في مثال (ج) متوسط الأجر الشهري للعمال في المصنعين I و II غير مختلف. فإن H_0 و H_1 يصاغا كما يلي:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

فإن منطقة رفض H_0 وقيمتها α والممثلة بالمساحة المظللة على المنحنى شكل (٧,٣) توضح كما يلي:



شكل (٧,٣)

(٧,٥) اتخاذ القرار الإحصائي

إذا وقعت قيمة الإحصاءة المحسوبة من العينة مثل الإحصاءات ... $Z, t, \chi^2, F,$ في منطقة رفض H_0 (المنطقة المظللة α) في الأشكال (٧,١)، (٧,٢) و (٧,٣) السابقة فإننا نرفض H_0 ونقبل الفرض البديل H_1.
أما إذا وقعت قيمة الإحصاءة في منطقة قبول H_0 (المنطقة غير المظللة $(1-\alpha)$)، في الأشكال السابقة (٧,١)، (٧,٢) و (٧,٣) فإننا لا نستطيع رفض فرض العدم H_0 ولكن نرفض الفرض البديل H_1.

(٧,٦) اختبار الفروض للمتوسط في العينات الكبيرة

لاختبار الفروض للمتوسط يصاغ فرض العدم H_0 والفرض البديل H_1 كما أوضحنا فيما سبق. ثم نختار الإحصاءة $Z = \dfrac{\overline{X} - \mu}{\sigma_{\bar{x}}}$ التي يكون لها توزيع يقترب من التوزيع الطبيعي القياسي $N(0,1)$، حيث الخطأ المعياري للمتوسط $\sigma_{\bar{x}}$ يساوي $\dfrac{\sigma}{\sqrt{n}}$ للمجتمعات الكبيرة ويساوي $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}$ للمجتمعات

الصغيرة وحجمها N عندما يكون نسبة حجم العينة n إلى حجم المجتمع N يزيد على 5%. يهمل المقدار $\sqrt{\frac{N-n}{N-1}}$ عندما تكون النسبة السابقة أقل من 5%. وتعتبر العينة كبيرة إذا كان حجم العينة n يزيد عن 30 مشاهدة. عندما يكون تباين المجتمع σ^2 مجهولا نستعيض بدلا منه بتباين العينة s^2. ونوضح طريقة اختبار الفروض للمتوسط بالمثال التالي.

مثال (١ , ٧)

اخذت عينة مكونة من 36 عجلا من مزرعة لتسمين الماشية فوجد أن متوسط الوزن للعينة هو $\overline{X} = 192$ كجم. اختبر الفرض القائل: أن متوسط وزن العجول بالمزرعة أقل من 200 كجم. إذا علم أن الانحراف المعياري للمجتمع $\sigma = 20$ كجم وذلك عند مستوى معنوية $\alpha = 0.05$.

الحل

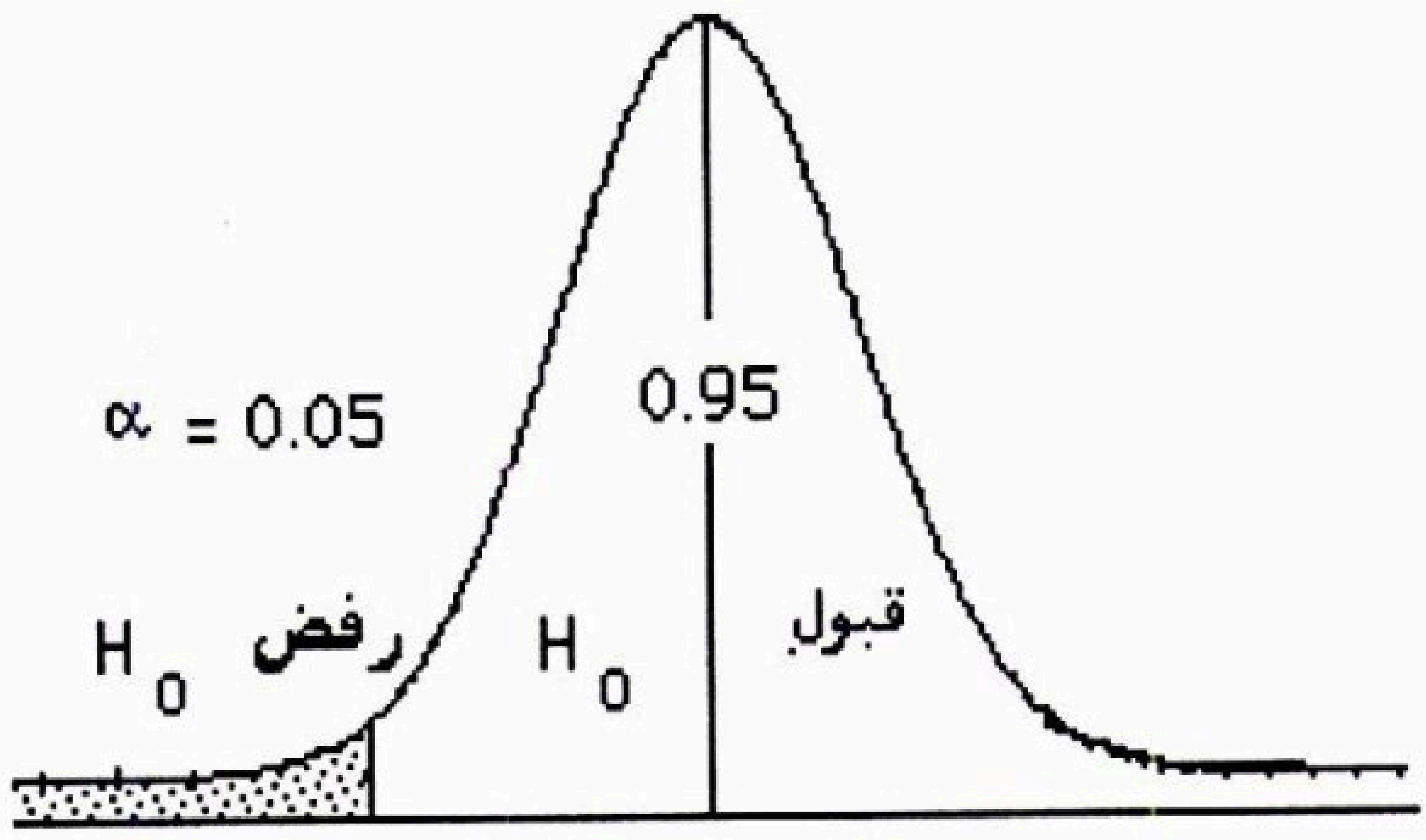
المعلومات التي لدينا في المثال هي $\overline{x} = 192$ كجم، $n = 36$، عجلا، $\sigma = 20$ كجم، فرض العدم $H_0 : \mu_0 = 200$ kg ويصاغ الحل في خطوات رئيسة كما يلي:

أولا: صياغة الفرض الإحصائي

$$H_0 : \mu_0 = 200 \text{ kg}$$

$$H_1 : \mu_1 < 200 \text{ kg}$$

أي أن الفرض البديل من طرف واحد من جهة اليسار بمستوى معنوية $\alpha = 0.05$ كما هو موضح بالرسم شكل (٤ , ٧) كما يلي:



شكل (٧,٤)

ثانيا: اختيار الإحصاءة ثم حسابها

نختار الإحصاءة Z_0 ثم نحسبها كما يلي:

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{192 - 200}{20 / \sqrt{36}} = -2.4$$

ثالثا: تحديد منطقة قبول ورفض H_0

نحدد فترة قبول H_0 من جدول (٧,٢) التالي للقيم الحرجة Z_0 وكما هي موضحة في شكل (٧,٤) السابق فإن فترة قبول H_0 هي $(-1.645, \infty)$ ومنطقة رفض H_0 خارج هذه الفترة وذلك عند $\alpha = 0.05$.

رابعا: اتخاذ القرار الإحصائي

وحيث $Z_0 = -2.4$ المحسوبة تقع خارج منطقة قبول H_0 فإننا نرفض H_0 القائل أن متوسط وزن العجول 200 كجم ونقبل الفرض البديل القائل أن الوزن أقل

من 200 كجم.

ملاحظة

يمكن تلخيص بعض القيم الحرجة للإحصاءة Z_0 عند بعض قيم مستويات المعنوية α وذلك في جدول (٧,٢) كما يلي:

جدول (٧,٢)

مستوى المعنوية α	0.1	0.05	0.025	0.01	0.005
قيم Z_0 الحرجة H_1 من طرف واحد	± 1.28	± 1.645	± 1.96	± 2.33	± 2.58
قيم Z_0 الحرجة H_1 من طرفين	± 1.645	± 1.96	—	± 2.58	—

(٧,٧) اختبار الفروض للمتوسط في العينات الصغيرة

إذا كانت العينة العشوائية المتوفرة لدينا صغيرة $(n < 30)$ لأسباب كثيرة منها زيادة التكلفة أوغيرها، وإذا كان المعلوم لنا أن المجتمع الإحصائي المأخوذة منه العينة طبيعيا$(\mu, \sigma^2)N$ فإنه يكون لدينا حالتين هما:

(ا) إذا كان تباين المجتمع σ^2 معلوما فإن الإحصاءة التي تستخدم لاختبار فرض العدم H_0 هي

$$Z_0 = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}}.$$

(ب) إذا كان تباين المجتمع σ^2 مجهولا فإن الإحصاءة التي تستخدم لاختبار فرض العدم H_0 هي $t_0 = \frac{\overline{x} - \mu}{s/\sqrt{n}}$. والتي تتبع توزيع t بدرجة حرية υ.

وحيث s هي الانحراف المعياري المحسوب من العينة ($\upsilon = n - 1$ درجات الحرية). ونوضح ذلك بالمثال التالي.

مثال (٢ , ٧)

اخذت عينة مكونة مـن16شابا بالغا وكان متوسط الطول لهـم158سم وكـان الانحراف المعياري للمجتمع معلوم ويساوي5 سم. اختبر الفرض القائل أن متوسط المجتمع الذي اخذت منه هذه العينة هو 160سم. وذلك عند مستوى معنوية $\alpha = 0.05$.

الحل

المعلومات من المثال (٢ , ٧) هي

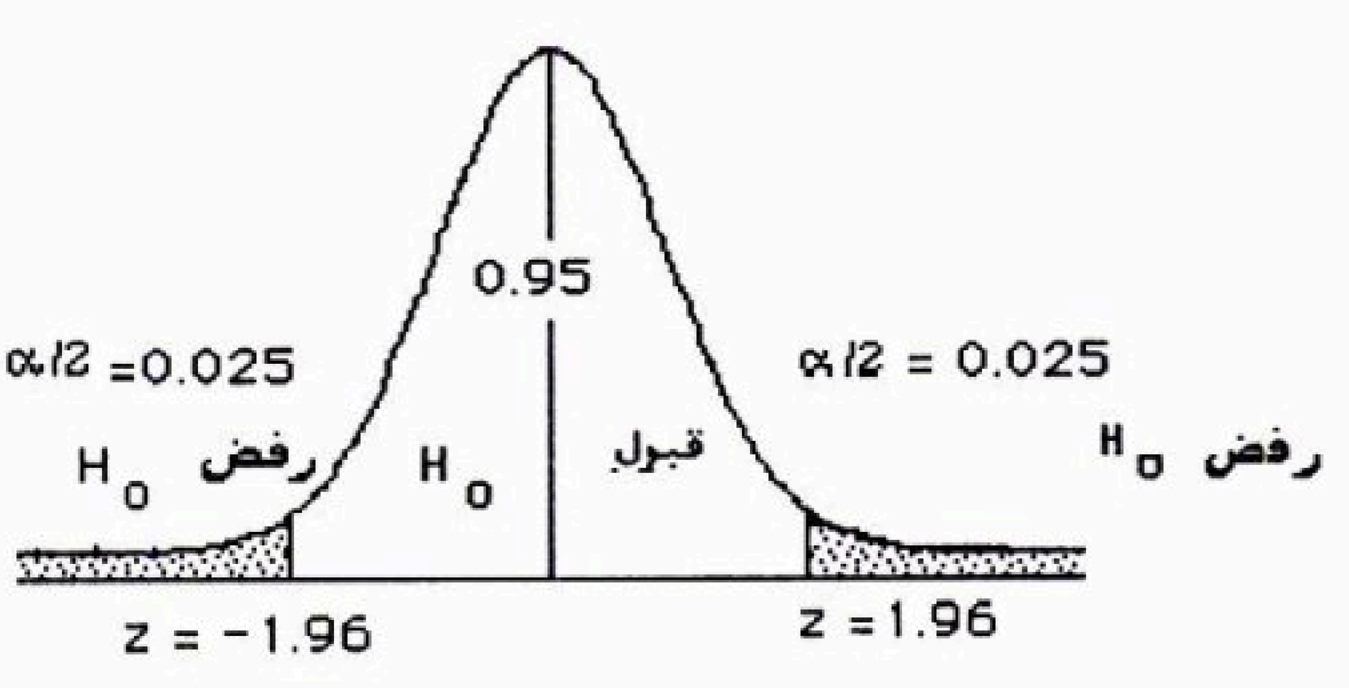
$$\overline{x} = 158 \text{ cm} , \quad \mu = 160 \text{ cm} , \ n = 16 , \quad \sigma = 5 \text{ cm}$$

أولا: صياغة الفرض الإحصائي:

$$H_0 : \mu_0 = 160 \text{ cm}$$

$$H_1 : \mu \neq \mu_0 \text{ cm}$$

أي أن الفرض البديل من طرفين بمستوى معنوية $\alpha = 0.05$ ويوضح بالجزء المظلل في شكل (٥ , ٧) كما يلي:



شكل (٥ , ٧)

ثانيا: اختيار الإحصاءة ثم حسابها

حيث إن حجم العينة صغيرا $n = 16$ والمجتمع المأخوذة منه العينة طبيعي والانحراف المعياري له s معلوما فإن الإحصاءة Z_0 تحسب كما يلي:

$$Z_0 = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \frac{158 - 160}{5/\sqrt{16}} = - 1.6$$

ثالثا: تحديد منطقة قبول ورفض H_0.

نحدد منطقة قبول H_0 بالفترة $(-1.96, 1.96)$ باستخدام جدول (٧,٢) للقيم Z_0 الحرجة من طرفين عند $\alpha = 0.05$.

رابعا: اتخاذ القرار الإحصائي

حيث $Z_0 = -1.6$ المحسوبة واقعة داخل فترة قبول H_0 وهي $(-1.96, 1.96)$ فإننا لا نستطيع رفض H_0 أي أن متوسط طول الأفراد البالغين في هذا المجتمع هو 160 سم بدرجة ثقة 95%.

مثال (٧,٣)

يدّعي مصنع لإنتاج حبال من الصلب تستخدم في المصاعد الكهربائية الخاصة بالمنازل بأن متوسط قوة قطع الحبال هي 6000 ثقل كجم اختيرت عينة مكونة من 9 حبال لحساب متوسط قوة قطعها وانحرافها المعياري فوجد أن المتوسط والانحراف المعياري للعينة هما $\overline{x} = 5750$ kg و $s = 250$ kg هل يمكن تأييد ادعاء المصنع عند مستوى معنوية $\alpha = 0.01$.

الحل

المعلومات في مثال (٧,٣) هي:

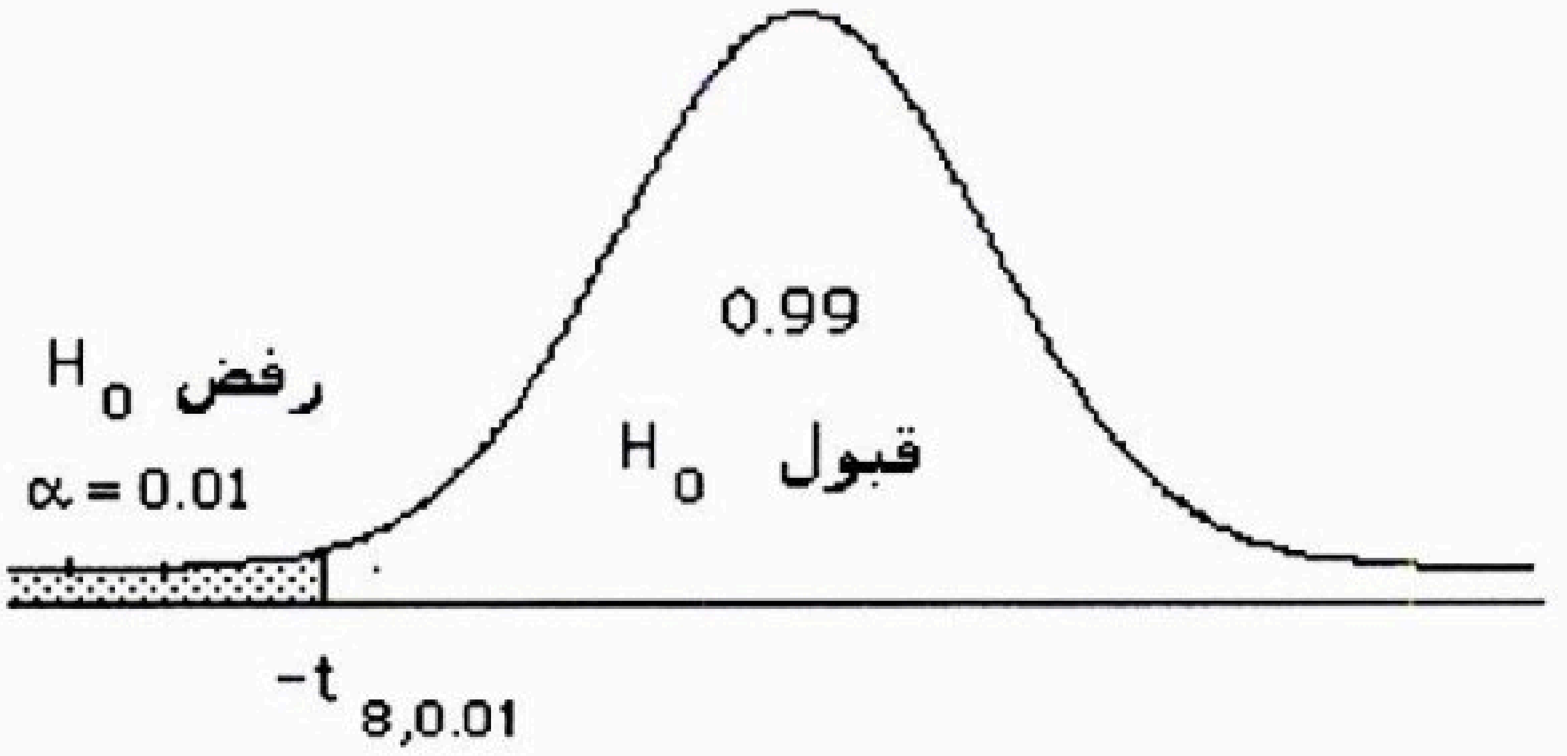
$$\overline{x} = 5750 \text{ kg}, s = 250 \text{ kg}, n = 9, \quad \mu_0 = 6000 \text{ kg}$$

أولا: صياغة الفرض الإحصائي هو.

$$H_0 : \mu_0 = 6000 \text{ kg}$$

$$H_1 : \mu < 6000 \text{ kg}$$

أي أن الفرض البديل هو من طرف واحد من جهة اليسار بمستوى معنوية $\alpha = 0.01$ وتوضح منطقة رفض H_0 بالجزء المظلل شكل (٦, ٧) كما يلي:



شكل (٦, ٧)

..

ثانيا: اختيار الإحصاءة ثم حسابها

نختار الإحصاءة t_0 حيث إن $\upsilon = n - 1 = 9 - 1 = 8$ فإن

$$t_0 = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{5750 - 6000}{250/\sqrt{9}} = -3$$

ثالثا: تحديد منطقة قبول ورفض H_0.

باستخدام جدول توزيع t رقم (٢) آخر الكتاب وكذلك شكل (٦, ٧) السابق فإن فترة قبول H_0 هي$(-t_{\upsilon,\alpha}, \infty)$ عند مستوى معنوية $\alpha = 0.01$ فتصبح

فترة قبـول H_0 هي $(-2.896, \infty)$ منطقـة رفـض H_0 خارج هذه الفترة وحيـث $\upsilon = 9 - 1 = 8$ وأن $t_{8,\ 0.01} = -2.896$.

رابعا: اتخاذ القرار الإحصائي.

حيث إن $t_0 = -3$ المحسوبة خارج منطقة قبول H_0 فإن القرار الإحصائي هو رفض H_0 أي أن ادعاء المصنع لحبال الصلب غير صحيح وأن العينة اعطت نتائج لمتوسط قوة القطع أقل من 6000 كجم بدرجة ثقة 99%.

(٧ , ٨) اختبار الفروض للنسبة في المجتمع

إذا كانت العينة العشوائية مأخوذة من مجتمع ينقسم إلى صفتين نـسبـة الصفة محل الدراسة هي المعلمة P (على سبيل المثال نسبة المدخنين بين الأشخاص البالغين في مدينة ما). ونرغب في اختبار الفروض حول هذه النسبة فإننا نختار عينة n $(n > 30)$ من المجتمع محل الدراسة ثم نحسب نسبة الصفـة من العينـة ولتكـن r (r نسبة المدخنين في العينة على سبيل المثال). ثم نحسب الإحصاءة

$$Z_0 = \frac{r - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

وذلك لاختبـار فرض العدم H_0 ضد الفرض البديـل H_1، ونوضح ذلك بمثال كما يلي:

مثال (٤ , ٧)

يدّعي مصنع للأدوية أن دواء من إنتاجه له فاعلية في التخفيف من الحساسية فترة عشرة ساعات بنسبة 90%. اخذت عينة من 200 شخص مصابين بالحساسية، أدى الدواء إلى تخفيف ألام 160 منهم. اختبر ما إذا كان ادعاء المصنع صحيح أم

غير صحيح عند مستوى معنوية $\alpha = 0.05$.

الحل

المعلومات من مثال (٧,٤) هي :

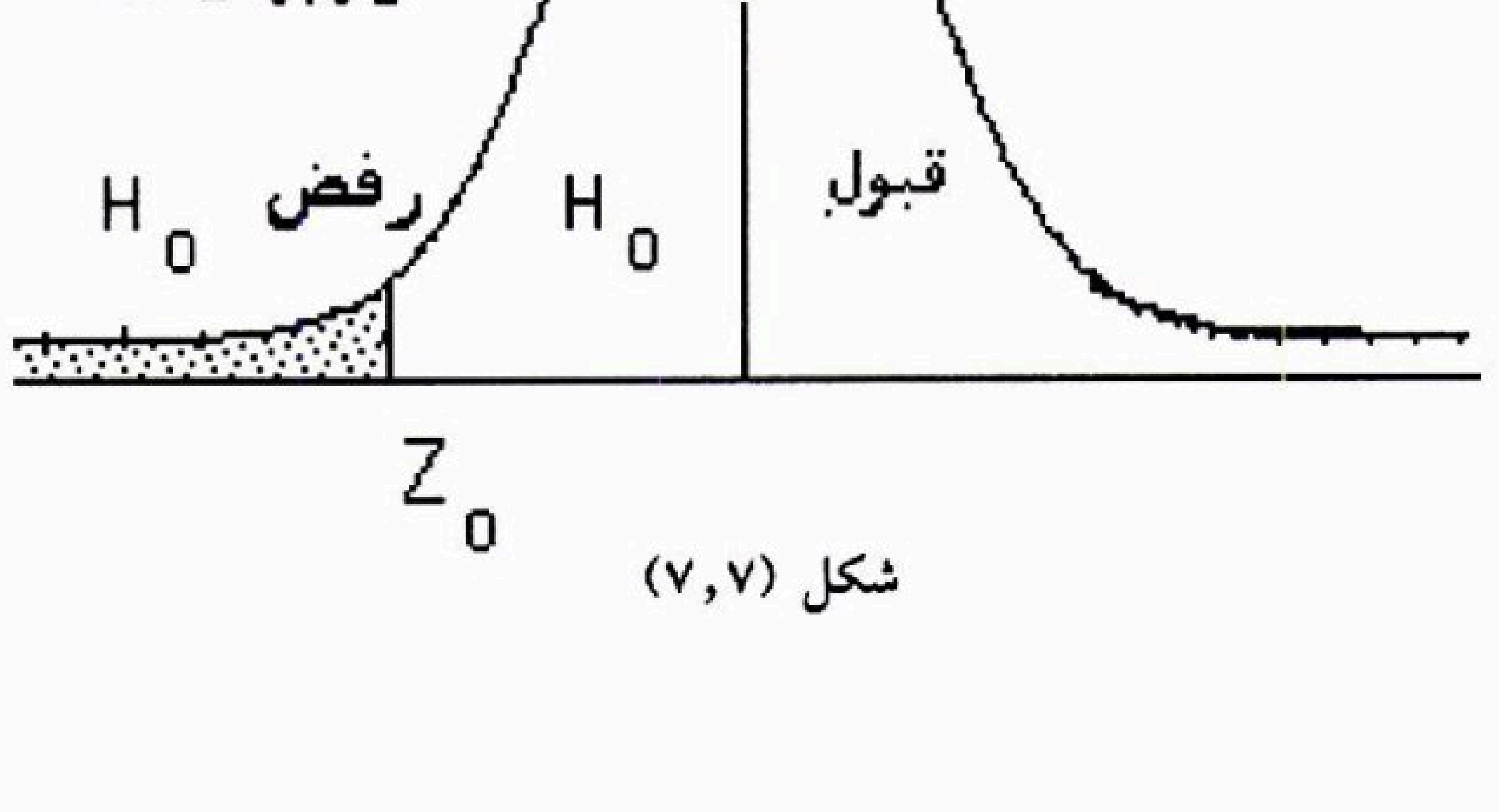
$$P_0 = 0.9\ ,\ r = \frac{160}{200} = 0.8\ ,\ n = 200$$

أولا: صياغة الفرض الإحصائي

$$H_0 : P_0 = 0.9$$

$$H_1 : P < 0.9$$

أي أن الفرض البديل هو من طرف واحد من جهة اليسار بمستوى معنوية $\alpha = 0.05$ وتوضح منطقة رفض H_0 بالجزئ المظلل شكل (٧,٧) كما يلي :



شكل (٧,٧)

ثانيا: اختيار الإحصاءة ثم حسابها.

نختار الإحصاءة Z_0 ويتم حسابها كما يلي:

$$Z_0 = \frac{r - P_0}{\sqrt{P_0(1-P_0)/n}} = \frac{0.8 - 0.9}{\sqrt{0.8(1-0.8)/200}} = -4.714$$

ثالثا: تحديد منطقة قبول ورفض H_0.

نستخدم جدول (٢ , ٧) لقيم Z_0 السابق نجد أن فترة قبول H_0 هي $(-1.645, \infty)$ عند مستوى معنوية $\alpha = 0.05$.

رابعا: اتخاذ القرار الإحصائي.

حيث قيمة $Z_0 = -4.714$ المحسوبة واقعة خارج منطقة قبول H_0 فإننا نرفض H_0 القائل أن نسبة الشفاء 90% ونقبل الفرض البديل القائل أن نسبة الشفاء أقل من 90% أي أن ادعاء المصنع غير صحيح بدرجة ثقة 95%.

(٩ , ٧) اختبار الفروض لتباين المجتمع

لاختبار الفرض حول تباين σ^2 للمجتمع. فإننا نحسب تباين العينة s^2. وبذلك يكون لدينا إحصاءة $\frac{(n-1)S^2}{\sigma^2}$ لها توزيع عيني يقترب من توزيع مربع كاي χ^2_υ حيث $\upsilon = n-1$ درجات الحرية ونوضح ذلك بمثال كما يلي

مثال (٥ , ٧)

اختيرت عينة عشوائية من 20 مفردة من مجتمع ما وكان التباين المحسوب من هذه العينة هو $s^2 = 266$ اختبر الفرض القائل أن تباين هذا المجتمع هو $\sigma^2 = 250$ بدرجة ثقة 0.99.

الحل

معلومات مثال (٧,٥) نلخصها كما يلي:

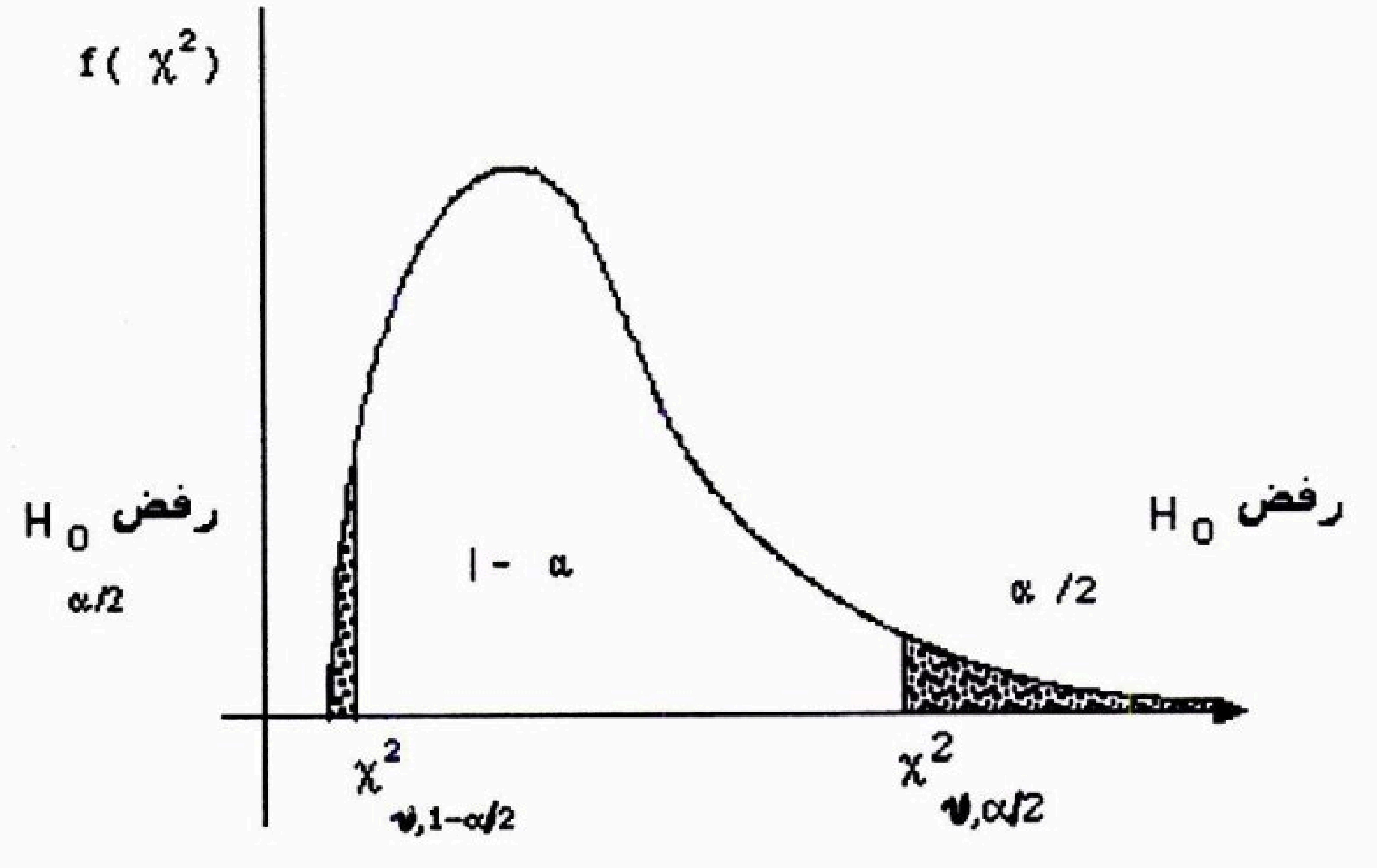
$$s^2 = 266, \ \sigma^2 = 250, \ n = 20, \ \upsilon = n\text{-}1 = 19$$

أولا: صياغة الفرض الإحصائي كما يلي.

$$H_0 : \sigma_0^2 = 250$$

$$H_1 : \sigma^2 \neq 250$$

أي أن الفرض البديل هو من طرفين بمستوى معنوية $\alpha = 0.01$ ونوضح منطقة رفض H_0 بالجزء المظلل في شكل (٧,٨) كما يلي:



شكل (٧,٨)

ثانيا: اختيار الإحصاءة ثم حسابها

نختار الإحصـــاءة $\chi_0^2 = \frac{(n-1)s^2}{\sigma^2}$ والتي لها توزيع χ_ν^2 وتحسب χ_0^2 كما يلي:

$$\chi_0^2 = \frac{(20-1)266}{250} = 20.216$$

ثالثا: تحديد منطقة قبول ورفض H_0.

باستخدام جدول (٣) لتوزيع χ_ν^2 آخر الكتاب ثم نحسب فترة قبول H_0 وهي($\chi_{19,0.005}^2$ ، $\chi_{19,0.995}^2$) فنجد أنها تساوي الفترة (6.644, 36.582) ومنطقة رفض H_0 خارج هذه الفترة. حيث $\nu = 20 - 1 = 19$ و حيث $\alpha = 0.01$.

رابعا: اتخاذ القرار الإحصائي.

قيمة $\chi_0^2 = 20.211$ المحسوبة واقعة داخل فترة قبول H_0 أي أننا لا نستطيع رفض H_0 وأن نرفض الفرض البديل القائل أن تباين المجتمع يختلف عن $\sigma^2 = 250$ عند مستوي معنوية $\alpha = 0.01$.

(١٠ , ٧) اختبار الفروض للفرق بين متوسطي مجتمعين

سنتناول في هذا البند اختبار المعنوية بين متوسطي مجتمعين في كل من الحالات التالية:

(ا) اختبار الفرضيات للفرق بين متوسطي مجتمعين مستقلين للعينات الكبيرة.

(ب) اختبار الفرضيات للفرق بين متوسطي مجتمعين مستقلين للعينات الصغيرة ومن مجتمعين طبيعيين.

(ج) اختبار الفرضيات للمتوسط للعينات المزدوجة.

وسنتناول كل حالة من الحالات السابقة بالشرح والأمثلة كما يلي:

(١ , ١٠ , ٧) اختبار الفروض للفرق بين متوسطي مجتمعين مستقلين للعينات الكبيرة

تستخدم الإحصـاءة $Z = \dfrac{(\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{n}}}$ والتي تقتـرب مـن

التوزيع الطبيعي القياسي $N(0,1)$ وفي حالة كون σ_2^2, σ_1^2 مجهولتين في المجتمعين نستعيض

عنهما بتباين العينتين s_2^2, s_1^2 لتصبح الإحصاءة $Z = \dfrac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ والتي

لها توزيع يقترب من $N(0,1)$ وذلك في حالة العينات الكبيرة.

مثال (٦ , ٧)

أعطي اختبار أحد المقررات لعينتين من شعبتين الأولى تتكون من 50 طالبا وكان متوسط درجاتها هو 64 درجة وانحراف معياري 6 درجات والثانية تتكـون من 60 طالبا وكان متوسط درجاتها هو 66 درجة وانحراف معيـاري 5 درجات. اختير الفرض القائل لا يوجد اختلاف في الأداء لهاتين الشعبتين عند مسـتـوى معنوية $\alpha = 0.05$.

الحل

معلومات مثال (٦ , ٧) نلخصها كما يلي:

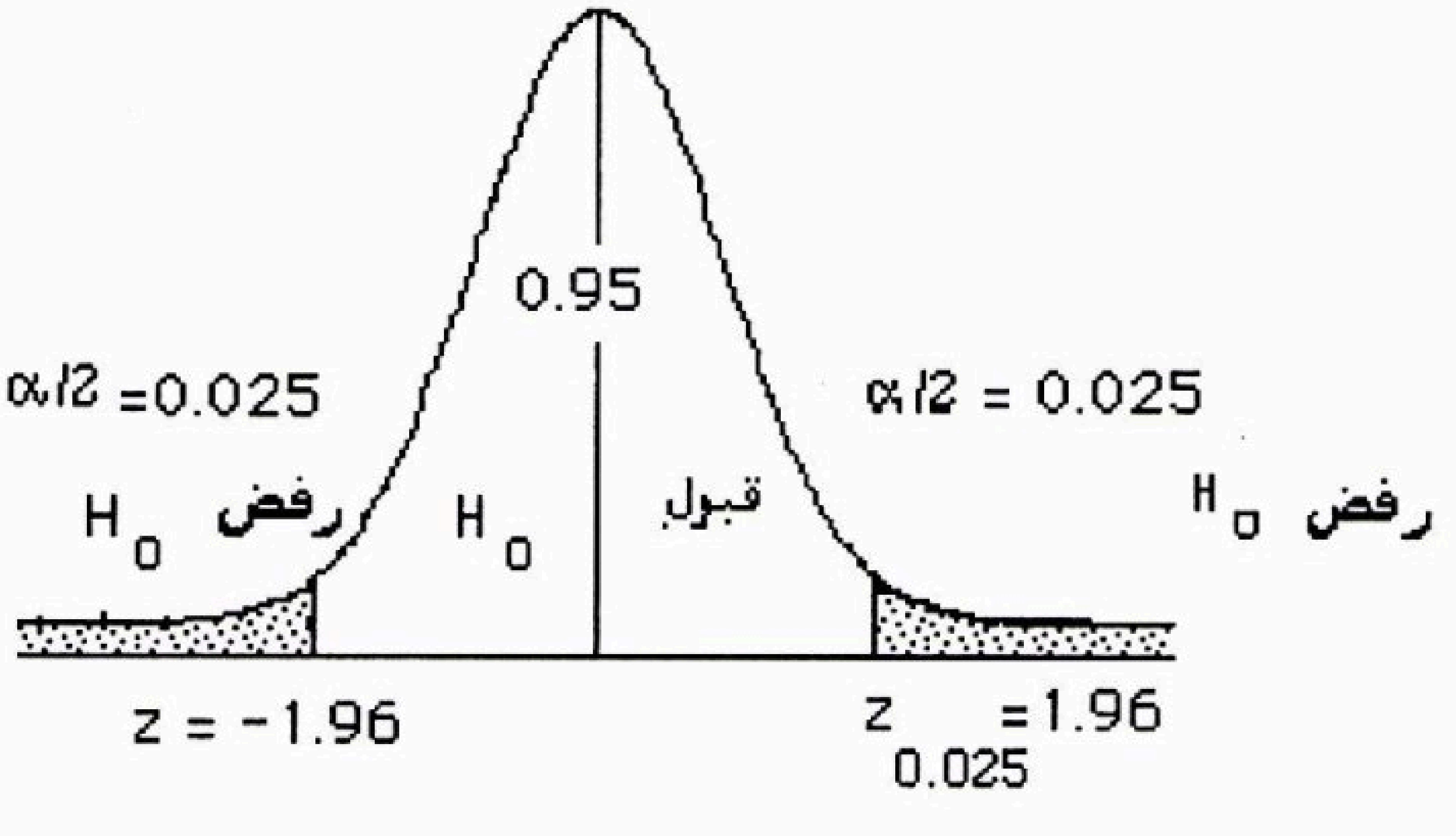
عينة الشعبة الثانية	**عينة الشعبة الأولى**
$\bar{x}_2 = 66$ درجة	$\bar{x}_1 = 64$ درجة
$s_2 = 5$ درجات	$s_1 = 6$ درجات
$n_2 = 60$ طالبا	$n_1 = 50$ طالبا

أولا: صياغة الفرض الإحصائي:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

أي أن الفرض البديل H_1 من طرفين بمستوى معنويـة $\alpha = 0.05$. ونوضح منطقـة رفض H_0 بالجزء المظلل في شكل (٧,٩) كما يلي:



شكل (٧,٩)

ثانيا: اختبار الإحصاءة ثم حسابها.

نختار الإحصاءة Z_0 وحسب صحة H_0 فإن $\mu_1 - \mu_2 = 0$ ثم نحسب الإحصاءة Z_0 كما يلي:

$$Z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(64 - 66) - 0}{\sqrt{\frac{(6)^2}{50} + \frac{(5)^2}{60}}} = -1.876$$

ثالثا: تحديد منطقة قبول ورفض H_0.

يستخدم جدول (٧ , ٢) لقيم Z_0 الحرجة فإن منطقة قبول H_0 عند مستوى معنوية $\alpha = 0.05$ من الطرفين $(-1.96, 1.96)$ ومنطقة رفض H_0 خارج هذه الفترة.

رابعا: اتخاذ القرار الإحصائي

قيمة $Z_0 = -1.876$ المحسوبة تقع داخل منطقة قبول H_0. أي أن لا نستطيع رفض H_0 ونرفض H_1 القائل أن أداء الشعبتين لهذا المقرر مختلف.

(٧ , ١٠ , ٢) اختبار الفروض للفرق بين متوسطي مجتمعين مستقلين من مجتمعين طبيعيين للعينات الصغيرة

(I): σ_2^2, σ_1^2 معلومتين ومن مجتمعين طبيعيين.

نختار الإحصاءة
$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$
لها توزيع الطبيعي القياسي $N(0,1)$ وحيث إن تباين المجتمعين σ_2^2, σ_1^2 معلومين وحجم كل من العينتين n_2, n_1 صغيرا أي أن $(n_1 < 30, n_2 < 30)$ فإنه يمكن استخدام الإحصاءة

Z السابقة في اختبار الفروض حول الفرق بين متوسطي مجتمعين مستـقـلـين. ونوضح ذلك من خلال المثال التالي.

مثال (٧,٧)

للمقارنة بين متوسطي أعمار سكان المدن والقرى حيث كان أعمار سكـان المدن يتبع توزيع طبيعي بانحراف معيـاري $\sigma_1 = 7$ سنوات وأعمار سكان القرى يتبع توزيع طبيعي بانحراف معياري 9 سنوات. أخذت عينة من سكان المدن $n_1 = 20$ شخصا كان متوسط أعمارهـم 63 سنة وكذلك عينة من سكان القـرى $n_2 = 25$ شخصا كان متوسط أعمارهم 60 سنة. اختبرالفرض القائل أنه لا يوجد اختلاف معنوي بين متوسطي أعمار المجتمعين عند مستوى معنوية .

الحل

نلخص معلومات مثال (٧,٧) كما يلي:

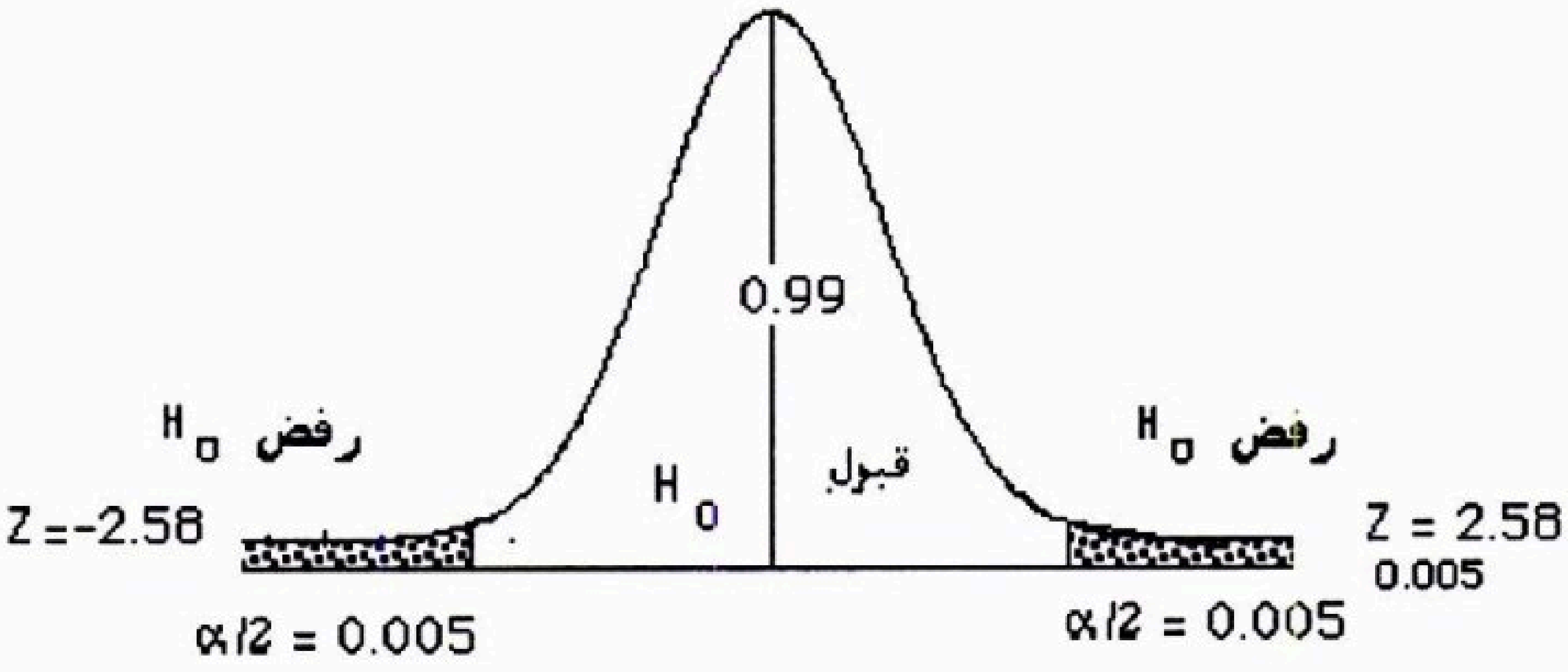
سكان القرى	سكان المدن
$n_2 = 25$ شخصا	$n_1 = 20$ شخصا
$\sigma_2 = 9$ سنوات	$\sigma_1 = 7$ سنوات
$\bar{x}_2 = 60$ سنة	$\bar{x}_1 = 63$ سنة

أولا: صياغة الفرض الإحصائي

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

أي أن الفرض البديـل H_1 من طرفين بمستوى معنويـة $\alpha = 0.01$ ونوضح منطقـة رفض H_0 بالجزء المظلل في شكل (٧,١٠) كما يلي:



شكل (٧,١٠)

ثانيًا: اختيار الإحصاءة ثم حسابها

حيث إن $\mu_1 - \mu_2 = 0$ تحت صحة فرض العدم H_0 ثم تحسب Z_0 كما يلي:

$$Z_0 = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{63 - 60}{\sqrt{\frac{(7)^2}{20} + \frac{(9)^2}{25}}} = \frac{3}{\sqrt{2.45 + 3.64}} = 1.2$$

ثالثا: تحديد منطقة قبول ورفض H_0.

يستخـــدم جــدول (٧,٢) لقيم Z_0 الحــرجة فإن منطقــة قبـــول H_0 هي (2.58, 2.58 -) ومنطقةرفض H_0 خارج هذه الفترة عند مستوى معنوية $\alpha = 0.01$ من الطرفين.

رابعا: اتخاذ القرار الإحصائي.

قيمة $Z_0 = 1.2$ المحسوبة واقعة في منطقة قبــول H_0 وبذلك لا نستطيــع

رفض H_0 . ونرفض الفرض البديل H_1 القائل أن متوسط أعمار سكان المدن يختلف عن متوسط أعمار سكان القرى بدرجة ثقة 99% .

(II) $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ومجهولة وحجم كل من العينتين n_1, n_2 صغيرا. نختار الإحصاءة

$$t_0 = \frac{(\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

لها توزيع t وتحسب درجات الحرية $\upsilon = n_1 + n_2 - 2$ ثم نحسب التباين s^2 المشترك لتباين العينتين s_1^2, s_2^2 بالعلاقة

$$s^2 = \frac{(n_1-1)s_1^2+(n_1-1)s_2^2}{n_1+n_2-2}$$

ثم نحسب الانحراف المعياري المشترك s بأخذ الجذر التربيعي للتباين s^2 ونوضح طريقة حساب الاختبار بالمثال التالي.

مثال (٨ , ٧)

في دراسة للمقارنة بين السعرات الحرارية الناتجة لنوعين من الفحم المنتج من منجمين مختلفين كانت النتائج التالية بملايين السعرات الحرارية هي :

المنجم الأول: 7930, 7860, 8380, 8230, 8400

المنجم الثاني: 7660, 8070, 7720, 7690, 7510

اختبر الفرض بأن المنجمين لهما نفس السعرات الحرارية عند مستوى معنوية $\alpha = 0.05$.

الحل

يمكن تلخيص المعلومات من مثال (٧,٨) كما يلي:

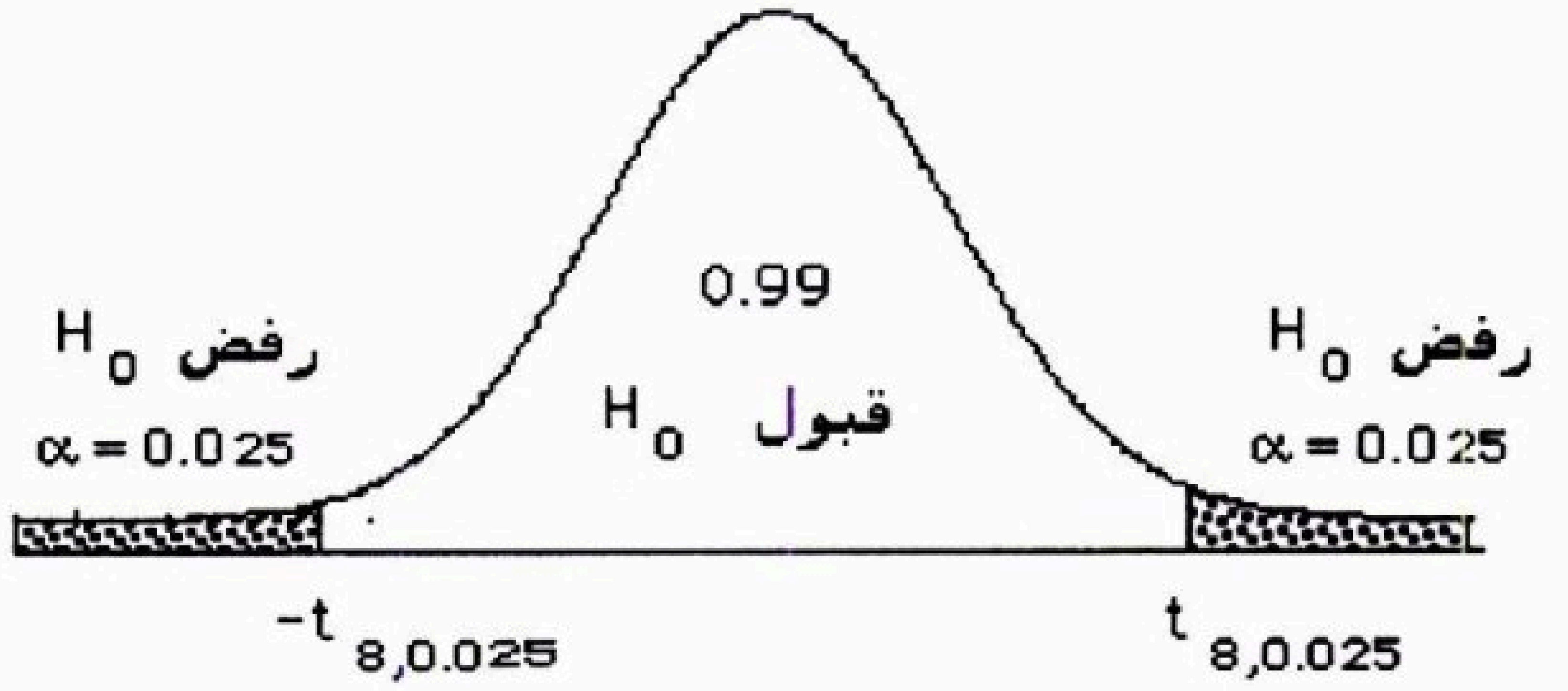
المنجم الثاني	**المنجم الأول**
$\bar{x}_2 = 7730$	$\bar{x}_1 = 8160$
$s_2 = 206.5$	$s_1 = 251.9$
$n_2 = 5$	$n_1 = 5$

أولا: صياغة الفرض الإحصائي .

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

أي أن الفرض البديل H_1 من طرفين بمستوى معنوية $\alpha = 0.05$ ونوضح منطقة رفض H_0 بالجزء المظلل في شكل (٧,١١) كما يلي :



شكل (٧,١١)

ثانيا: اختيار الاحصاءة ثم حسابها

نختار الإحصاءة t_0 وتحت صحة H_0 فإن $\mu_1 - \mu_2 = 0$ وأن لها توزيع t_ν بدرجات

الحرية $\upsilon = n_1 + n_2 - 2$ حيث قيمتها هي $\upsilon = 5 + 5 - 2 = 8$ ثم نحسب t_0 كما يلي:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{8160 - 7730}{\sqrt{\frac{253800 + 170600}{5 + 5 - 2}}\sqrt{\frac{1}{5} + \frac{1}{5}}} = 2.95$$

ثالثا: تحديد منطقة قبول ورفض H_0 .

من جدول (٢) لتوزيع t آخر الكتاب نجد أن فترة قبول H_0 هي $(-t_{8,0.025}\ ,\ t_{8,0.025})$ وقيمتها هي الفترة $(-2.306\ ,\ 2.306)$ ومنطقة رفض H_0 خارج هذه الفترة .

رابعا : القرار الإحصائي.

قيمة $t_0 = 2.95$ المحسوبة خارج منطقة قبول H_0 أي أن نرفض H_0 ونقبل الفرض البديل H_1 القائل أن متوسط السعرات الحرارية مختلف للمنجمين.

(٣ ,١٠ ,٧) اختبار الفرضيات للمتوسط للعينات المزدوجة

استخدمنا في الاختبارات السابقة اختبار الفرق بين متوسطي عينتين مستقلتين بمعنى أن العينة الأولى مفرداتها مستقلة عن مفردات العينة الثانية . ولكن قد يحدث في الحياة العملية أن القراءات تكون في صورة أزواج مرتبة (x,y) . فمثلا إذا أخذنا قراءات عينة عشوائية في المرة الأولى وكانت القراءات لها هي $x_1, x_2, \ldots, x_n$ ثم وضعنا هذه العينة تحت تأثير مؤثر نرغب في دراسة تأثيره على مفردات هذه العينة بعد فترة زمنية محددة . وفي المرة الثانية كانت القراءات لها هي $y_1, y_2, \ldots, y_n$. وإجراء اختبار الفروض لمتوسط الفرق بين القراءتين للعينة في المرتين الأولى والثانية حيث نحصل على متغير جديد D يأخذ القيم $d_1, d_2, \ldots, d_n$ حيث يتم حسابه كما يلي :

$$d_i = x_i - y_i \;, \quad i = 1,2 ,..., n$$

ويتم حساب متوسط فروق القراءات d_i للعينة المزدوجة $\overline{d}$ والانحراف المعياري S_d وتستخدم إحصاءة جديدة هي $\frac{\overline{D} - \mu_d}{S_{\overline{d}}}$ يكون لها توزيع معاينة يتبع توزيع t بدرجات الحرية υ (حيث $\upsilon = n - 1$) والخطأ المعياري لها $S_{\overline{d}}$ يحسب من العلاقة $S_{\overline{d}} = \frac{S_d}{\sqrt{n}}$ ونوضح ذلك بالمثال التالي .

مثال (٩ , ٧)

إذا كان لدينا درجات مجموعة مكونة من سبعة طلاب في مادتي الكيمياء والفيزياء بيانها كما يلي :

درجات الكيمياء X : 62, 82, 77, 57, 62, 90, 82

درجات الفيزياء Y : 53, 75, 65, 55, 67, 85, 79

والمطلوب استخدام اختبار الفرض القائل أنه لا يوجد فرق بين متوسطي درجات الكيمياء والفيزياء عند مستوى معنوية $\alpha = 0.01$.

الحل

نحسب المتوسط $\overline{d}$ ، s_d للفروق $d = x - y$ وهي من معلومات مثال (٩ , ٧) فإن قراءات d هي

9, 7, 12, 2, - 5, 5, 3

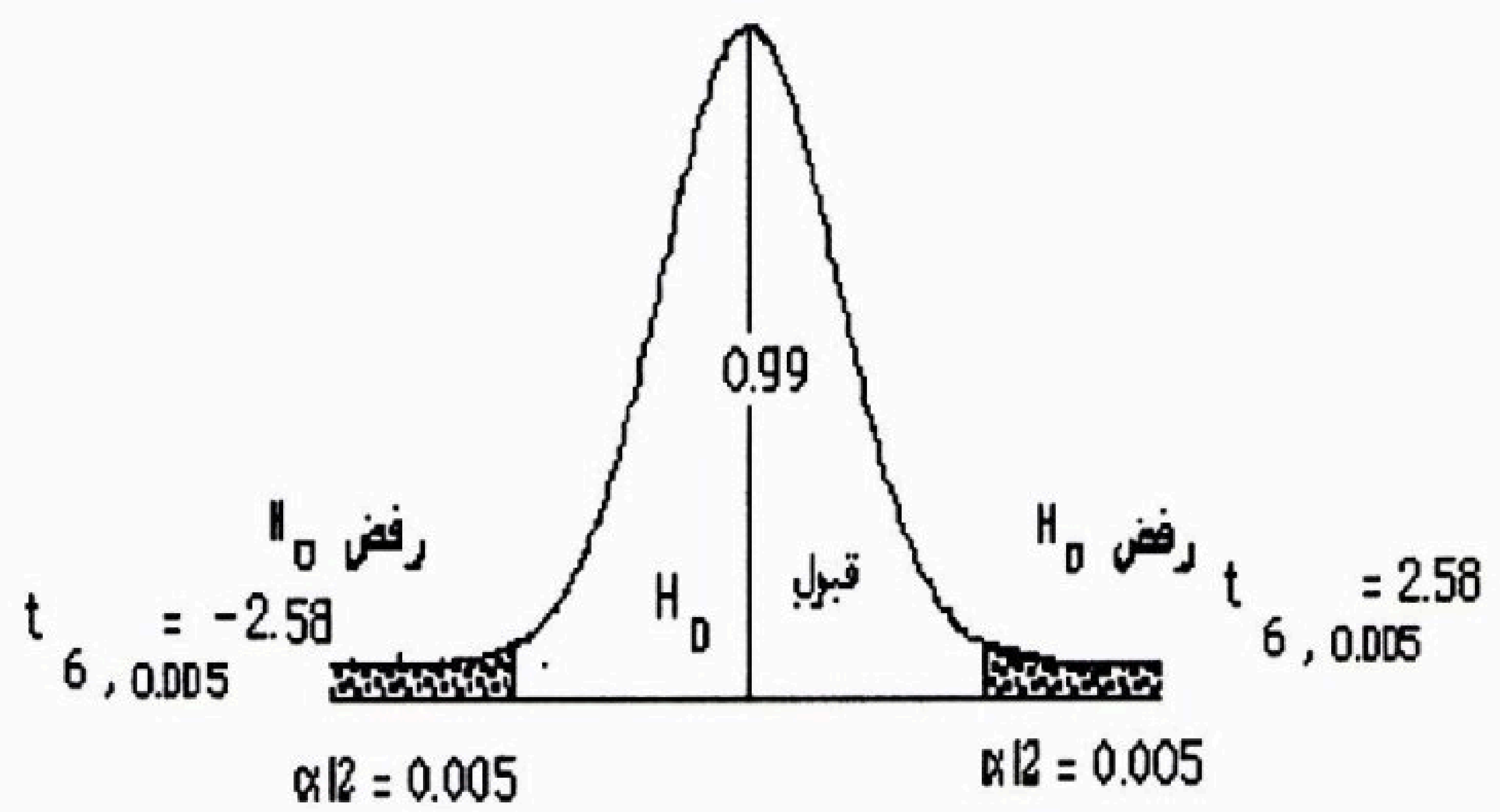
فإن متوسط الفروق هو $5 = \overline{d}$ والانحراف المعياري $s_d = 5.686$ ونتابع خطوات اختبار الفروض كما يلي :

أولا: صياغة الفرض الإحصائي.

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d \neq 0$$

أي أن الفرض البديل H_1 من طرفين بمستوى معنوية $\alpha = 0.01$ ونوضح منطقة رفض H_0 بالجزء المظلل في شكل (٧,١٢) كما يلي :



شكل (٧,١٢)

ثانيا: اختيار الإحصاءة ثم حسابها .

نختار الإحصاءة t_0 وتحت صحة H_0 فإن $\mu_d = 0$ ودرجات الحرية $\upsilon = n-1$ وقيمتها هي $\upsilon = 7-1 = 6$ ثم نحسب الإحصاءة t_0 كما يلي:

$$t_0 = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{5}{5.686 / \sqrt{7}} = 2.327$$

ثالثا : تحديد منطقة قبول ورفض H_0 .

نستخدم جدول (٢) لتوزيع t آخر الكتاب فنجد أن فتر قبول H_0 هي$(-t_{6,0.005} \;,\; t_{6,0.005})$ وقيمتها $(-3.707, 3.707)$ عند مستوى معنوية $\alpha = 0.01$ ونرفض H_0 خارج هذه الفترة .

رابعا : القرار الإحصائي .

قيمة $t_o = 2.327$ المحسوبة واقعة في منطقة قبول H_0 وبذلك لا نستطيع رفض H_0 ونرفض الفرض البديل H_1 القائل أن متوسط درجات الكيمياء لايختلف عن متوسط درجات الفيزياء بدرجة ثقة 99%.

(١١ , ٧) اختبار الفرضيات للفرق بين نسبتي مجتمعين

اختبار الفرق بين نسبتين لعينتين عشوائيتين مستقلتين للعينات الكبيرة ، لمعرفة ما إذا كان الفرق يرجع إلى عامل الصدفة أم أن هناك فروقا حقيقية ؟ مثل التدخين لعينتين من الأشخاص البالغين في بلدين مختلفين، ونسبة الشفاء من مرض معين باستخدام عقارين مختلفين لعينتين من المرضى . . . الخ . فإذا كانت النسبتين r_1 , r_2 من عينتين عشوائيتين حجمها على الترتيب n_1 , n_2 فإن توزيع المعاينة للفرق $r_1 - r_2$ يقترب من التوزيع الطبيعي فإنه يمكن صياغة إحصاءة Z تعطى بالعلاقة

$$Z = \frac{(r_1 - r_2) - (P_1 - P_2)}{\sqrt{\dfrac{P_1(1-P_1)}{n_1} + \dfrac{P_2(1-P_2)}{n_2}}}$$

تقترب من التوزيع الطبيعي القياسي N(0,1) وعندما تكون نسبتي المجتمعين P_1 , P_2 مجهولتين فإن الإحصاءة السابقة تصبح

$$Z = \frac{(r_1 - r_2) - (P_1 - P_2)}{\sqrt{r(1-r)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

حيث r المشتركة تعطى بالعلاقة $r = \frac{n_1 r_1 + n_2 r_2}{n_1 + n_2}$ ونوضح ذلك بالمثال التالي.

مثال (٧, ١٠)

مجموعتان A,B ، تتكون كل منهما من 100 شخص مصابين بمرض معين. أعطي مصل للمجموعة A ولـم يعط للمجموعة B (التي تسمى بالمجموعة الضابطة). بخلاف ذلك فإن المجموعتين تعاملان معاملة متماثلة. وقد وجد أنه في المجموعة A شفي 75 شخصا من المرض بينما في المجموعـة B شفي 65 شخصا . اختبـر الفــرض أن المصـل يساعــد على الشفــاء من المـرض باستخـدام مستوى معنوية $\alpha = 0.01$.

الحل

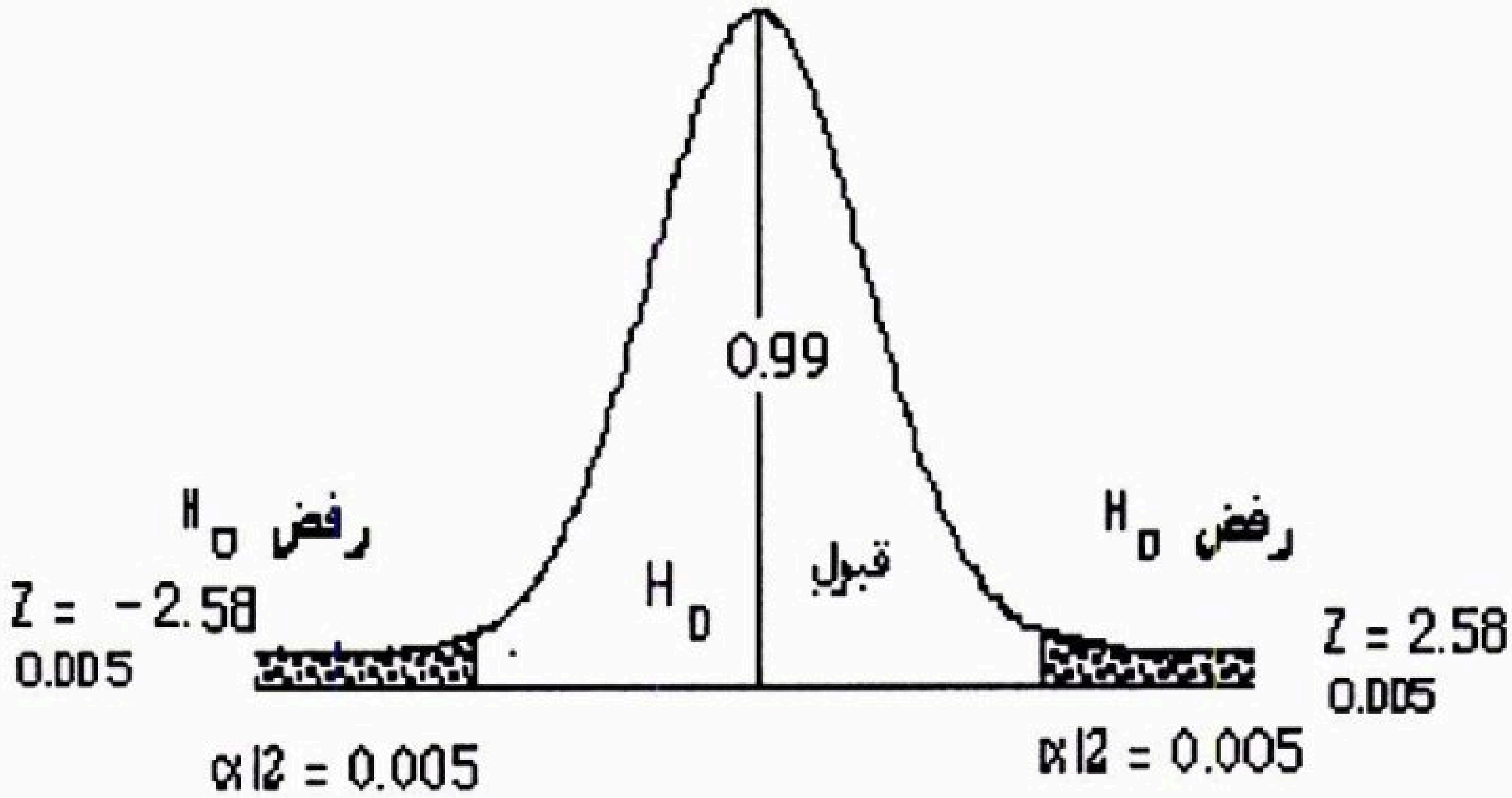
يمكن تلخيص معلومات مثال (٧, ١٠) كما يلي :

المجموعة B	المجموعة A
$r_2 = 0.65$	$r_1 = 0.75$
$n_2 = 100$	$n_1 = 100$

أولا : صياغة الفرض الإحصائي .

$$H_0 : P_1 - P_2 = 0$$
$$H_1 : P_1 \neq P_2$$

أي أن الفرض البديل من طرفين بمستوى معنويـة $\alpha = 0.01$ ونوضح منطقة الرفض بالجزء المظلل في شكل (٧, ١٣) كما يلي.



شكل (١٣, ٧)

ثانيا : اختيار الإحصاءة ثم حسابها .

نختار الإحصــاءة Z_0 وتحت صحة الفرض H_0 فإن $P_1 - P_2 = 0$ ثم يتــم حسابها كما يلي :

$$Z_0 = \frac{r_1 - r_2 - 0}{\sqrt{r(1-r)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.75 - 0.65}{\sqrt{0.7 \times 0.3\left(\frac{1}{100} + \frac{1}{100}\right)}} = 1.54$$

حيث تم حساب r وتمثل النسبة المشتركة من نسبتي العينتين r_1, r_2 كما يلي :

$$r = \frac{n_1 r_1 + n_2 r_2}{n_1 + n_2} = \frac{0.75(100) + 0.65(100)}{100 + 100} = 0.7$$

ثالثا: تحديد منطقة قبول ورفض H_0 .

يستخدم جدول (٧, ٢) لقيم Z_0 الحرجة فإن فترة قبول H_0 تصبــح الفـترة (2.58 ,2.58 -) ومنطقة رفض H_0 خارج هذه الفترة عند $\alpha = 0.01$.

رابعا : القرار الإحصائي .

قيمة $Z_0 = 1.54$ المحسوبة واقعة في منطقة قبول H_0 وبذلك نرفض الفرض البديل H_1 القائل أن المصل له تأثير في الشفاء عند مستوى معنوية $\alpha = 0.01$.

(١٢ , ٧) تمارين

١ - يدّعي صانع سنارات سمك أن اختباره للسنارات ستـعـطـي 8 رطل عنـد الاختبار . فهل هو محق في دعواه ؟ إذا كانت عينة حجمها 50 تعطي $\overline{x} = 7$ أرطال و $s = 1.4$ رطل عند مستوى معنوية

$\alpha = 0.01$(i) $\alpha = 0.05$ (ii) .

٢ - طبيب أسنان وجد أن 6 من مرضاه يحتاجون إلى 2,3,6,0,4,3 عمليات حشو اختبر الفرض القائل متوسط عمليات الحشو للمرضى يساوي $\mu = 5$ عمليات عند مستوى معنوية $\alpha = 0.05$.

٣ - إن كانت قيمة $\overline{x} = 82$,$\sigma = 16$, $n = 100$ اختبر الفرض بـأن $\mu = 86$ عند مستوى معنوية $\alpha = 0.01$.

٤ - خبرة سنوات عدة في امتحان اللغة الإنجليزية لدخول الجامعة أعطى مستوى النقاط 64 نقطة بانحراف معياري 8 درجات. حصل طلاب مدينة ما وعددهم 45 على متوسط نقاط قـدره 68 نقطة . فهل يمكن التحقق مـن أن هـؤلاء الطلاب أحسن مستوى في الإنجليزية من بقية الطلاب ؟ عند مستوى معنوية $\alpha = 0.05$.

٥ - لمعرفة أثر غذاء معين على زيادة الوزن أخذت عينة من خمسة فـئـران وتم تغذيتها بهذا لغذاء وكانت أوزانها بعد التغذية

2.4, 2.3, 1.5, 1.4, 1.6

هل تستطيع أن تحكم على أن هذه العينة مأخوذة من مجتمع متوسط الوزن فيه هو 1.8 أم لا ؟ عند مستوى معنوية $\alpha = 0.05$.

٦ - نسبة الطلاب الذين حصلوا على تقدير A في مادة الرياضيات في إحدى الجامعات خلال فترة طويلة من الزمن كانت 10% ، خلال فصل دراسي معين حصل 40 طالبا على تقدير A من بين مجموعة 300 طالب . اختبر معنوية هذه النتيجة عند مستوى معنوية $\alpha = 0.01$.

٧ - عينة عشوائية من 300 مسمار من إنتاج ماكينة (I) و 200 مسمار من إنتاج ماكينة (II) وجد من إنتاج (I) 28 مسمار تالفا ووجد من إنتاج (II) 15 مسمارتالفا .

اختبر الفرض القائل أن :

(i) هناك اختلاف في أداء الماكينتين .

(ii) الماكينة (II) تعمل بصورة أفضل من الماكينة (I) .

٨ - من بين 400 نبات من نباتات القمح تبين أن 70 نباتا تتوافر فيها صفة معينة. وطبقا لقانون الوراثة لمندل فإن 25% من محصول القمح يجب أن تتوافر فيه هذه الصفة. فهل تتفق هذه النتيجة مع قانون مندل للوراثة ؟ استخدم مستوى معنوية $\alpha = 0.01$.

٩ - إذا كان من المعروف أن الضغط الداخلي لكرات التنس المنتجة بواسطة أحد المصانع يتبع توزيعا طبيعيا متوسطه 28 رطلا لكل بوصة مربعة وتباين 0.25. وقد بدأ هذا المصنع في اتباع طريقة جديدة لإنتاج هذه الكرات تحافظ على المتوسط كما هو 28 رطلا ولكن على أمل أن تقلل من قيمة التباين . وبعد بدء الإنتاج بالطريقة الجديدة اختيرت عينة عشوائية من 25 كرة فكان التباين لها $s^2 = 0.15$ فهل تدل هذه النتيجة أن المصنع حقق تقليل التباين؟ استخدم معنوية $\alpha = 0.05$.

١٠ - لاختبار الفروض أن طلاب الجامعة أطول من طالباتها فتم اختيار مجموعتين من الطلاب والطالبات A, B وكانت البيانات كما يلي:

مجموعة الطالبات B	مجموعة الطلاب A
$\bar{x}_2 = 168$ cm	$\bar{x}_1 = 170$ cm
$s_2 = 10$ cm	$s_1 = 12$ cm
$n_2 = 40$ طالبة	$n_1 = 50$ طالب

فهل يمكن القول بأن متوسط أطوال طلاب الجامعة أكبر من متوسط أطـوال طالباتها ؟ استخدم مستوى معنوية $\alpha = 0.05$.

١١ - إذا كانت أوزان الأفراد الذكور تتبع توزيعا طبيعيا للمقارنة بين مجموعتين A, B لفئتي العمر (50 – 40) و(30 – 20) كانت نتائج الوزن كما يلي :

مجموعة B	مجموعة A
فئة العمر (50 – 40)	فئة العمر (30 – 20)
$\bar{x}_2 = 68.6$ kg	$\bar{x}_1 = 70.19$ kg
$s_2 = 12.56$ kg	$s_1 = 8.7$ kg
$n_2 = 15$ شخصا	$n_1 = 10$ شخصا

فهل تدل هذه البيانات على أن الأفراد الذكور في فئة العمر (30–20) أثقل وزنا من الافراد في فئة العمر (50–40) . استخدم مستوى معنوية 0.01 .

١٢ - يقوم أحد خبراء التغذية بتجربة نظام جديد للتغذية لتخـفـيـض الـوزن . فاختارعشرة أشخاصا عشوائيا وسجل أوزانـهـم (X) ثم طبق نظام التغذيـة الجديد على هؤلاء الأشخاص لمدة شهر واحد ثم سجل اوزانهم (Y) في نهاية الفترة فحصل على النتائج الآتية :

رقم الشخص	1	2	3	4	5	6	7	8	9	10
X الوزن في البداية	81	64	67	72.5	69	70	79	82	36	61
Y الوزن في النهاية	79	64.5	64.5	72	67	72	77	82	62	60

فهل تدل هذه القراءات على أن طريقة التغذية الجديدة تؤدي إلى انخفاض في وزن الأشخاص الذين يتبعونها ؟

استخدم مستوى معنوية $\alpha = 0.05$. وافتراض أن أوزان الأشخاص قبل وبعد اتباع نظام التغذية تتبع توزيعا طبيعيا .

١٣ - في بحث أجري عن التدخين عام 1980م اختيرت عينة من 800 شخص من سكان إحدى المدن فكان عدد المدخنين من بيهم 200 شخص ومنذ ذلك الوقت دعمت الحملات التي تدعو إلى الامتناع عن التدخين في 1985. اختيرت عينة عشوائية حجمها 1000 شخص من سكان المدينة نفسها فكان من بينها 210 مدخن . فهل تدل هذه النتائج على أن نسبة المدخنين قد انخفضت ؟ استخدم مستوى معنوية $\alpha = 0.05$.

١٤ - بإفتراض أن درجات كلا من الطلبة والطالبات في مقرر من مقررات الإحصاء بجامعة الملك سعود يتبع التوزيع الطبيعي . اختيرت عينة عشوائية من كل من الطلاب والطالبات من دارسي هذا المقرر لأحد الفصول الدراسية وكانت النتائج كما هي في الجدول التالي :

	حجم العينة	متوسط الدرجات	الانحراف المعياري
الطلاب	10	68	7
الطالبات	8	75	5

اختبر الفروض التالية وذلك لمستوى معنوية $\alpha = 0.05$.

(i) أن متوسط درجات الطلاب قد زاد عما كان عليه الفصل السابق لهذا المقرر وهو 65.

(ii) أن درجات الطلاب أصبحت أكثر تجانسا من درجات الطالبات

(iii) أن متوسط درجات الطلاب أقل من متوسط درجات الطالبات.

١٥- مصنع لإنتاج نوع معين من الأجهزةكانت نسبة المعيب من إنتاجه هــو 0.03 أجريت له صيانة فكان إنتاج اليوم الأول بعد الصيــانــة 500 جهاز منها 12 معيبة فهل تدل تلك النتائج على إنخفاض نسبة المعيب بعد الصيانة . وإذا تم تركيب مصنع جديد وأخذ إنتاج يوم فكان 600 جهاز منها 12معيبة فهل تدل النتائج على أن نسبة المعيب من إنتاج المصنع بعد الصيانة أعلى نسبة من نسبة المعيب في حالة المصنع الجديد عند مستوى معنوية $\alpha = 0.01$.

الفصل الثامن

اختبارات مربع كاي
Chi-square Tests

● مقدمة ● اختبارات مربع كاي لجودة التوفيـق ● اختبارات مربع كاي ● اختبارات مربع كـاي للاستقلال والتجانس ● تمارين

(٨,١) مقدمة
Introduction

سبق أن درسنا في الفصل السابق اختبـار Z للمتوسطات وللنسب في العينـات الكبيرة وكذلك اختبـار t للمتوسطات في العينات الصغيرة، وكل من اختبـار Z واختبار t قد لا يساعد إلى درجة كبيرة في اختبار العلاقة بين ما هو متاح للباحث من معلومات مشاهدة معينة، وما يمكن حسابه من معلومات (نظرية) من خـلال قانون رياضي معين يحكم ويحرك هذه الظاهرة ويحدد مسلكها العام. وهنا يلجأ الباحث إلى استخدام توزيع مربع كاي χ^2 لإجراء مثل هذا الاختبار لمعرفة معنوية العلاقة بين البيانات المشاهدة والبيانات المتوقعة من خلال قانون رياضي معروف.

وتتسع المجالات المختلفة لإجراء اختبار χ^2 لتشمل معالجة العلاقة المعنوية بين البيانات الوصفية مثل صفة التدخين والتعليم ولون البشرة والعيون والشعر. . . الخ، والبيانات الكمية مثل الطول والوزن ودرجات الامتحان والذكاء. . . الخ، والتي يمكن صياغة هذه الصفات وما يناظرها من البيانات التي تقاس كميا في أحد جداول التوافق أو الاقتران. فمثلا يمكن دراسة العلاقة المعنوية نتيجة إعطاء مصل

معين ضد الإصابة بأحد الأمراض، حيث يتم تقسيم مفردات العينة إلى (مجموعة أخذت المصل وأخرى لم تأخذ المصل) والنتيجة أيضا تقسم إلى (مجموعة لم تصب بالمرض وأخرى أصيبت بالمرض) وهكذا لكثير من التطبيقات في المجالات المختلفة. وربما يكون أهم مجالات اختبار χ^2 تطبيقا هو علم الوراثة والزراعة والفيزياء. . . الخ، حيث يتم اختبار صحة بعض النظريات (في الوراثة مثلا) من خلال البيانات المشاهدة المتاحة على ظاهرة تخضع لقانون وراثي محدد.

ومن ضمن المجالات المختلفة لاختبار χ^2 استخدامه في اختبارجودة التوفيق لبعض المنحنيات مثل منحنى الطبيعي ومنحنى توزيع ذي الحدين ومنحنى بواسون. . . الخ، أي أن اختبار مدى تبعية بعض المشاهدات المتاحة لتوزيع احتمالي معين له دالة توزيع احتمالي معروفة. كما يستخدم اختبار χ^2 في اختبار مدى دقة معادلة معينة إلى جانب استخدامه في اختبار معنوية التباين وايجاد فترات الثقه له.

وخلاصة القول فإن اختبارات χ^2 تتركز فيما يلي:

١ - اختبارات جودة التوفيق.

٢ - اختبارات التجانس والاستقلال في جداول التوافق والاقتران.

وسنتناول ذلك بالشرح والتفصيل فيما يلي.

(٢, ٨) اختبارات مربع كاي لجودة التوفيق

Chi-square Goodness-of-Fit Test

تهدف اختبارات χ^2 لجودة التوفيق إلى اختبار فرض العدم H_0 القائل إن مجموعة المشاهدات التي يتم اختبارها تتبع توزيع احتمالي معين أو نظرية معينة، ضد الفرض البديل H_1 القائل إن هذه المشاهدات لا تتفق مع هذا التوزيع أو هذه النظرية وبفرض صحة فرض العدم H_0 فإن الإحصاءة (١, ٨) التالية:

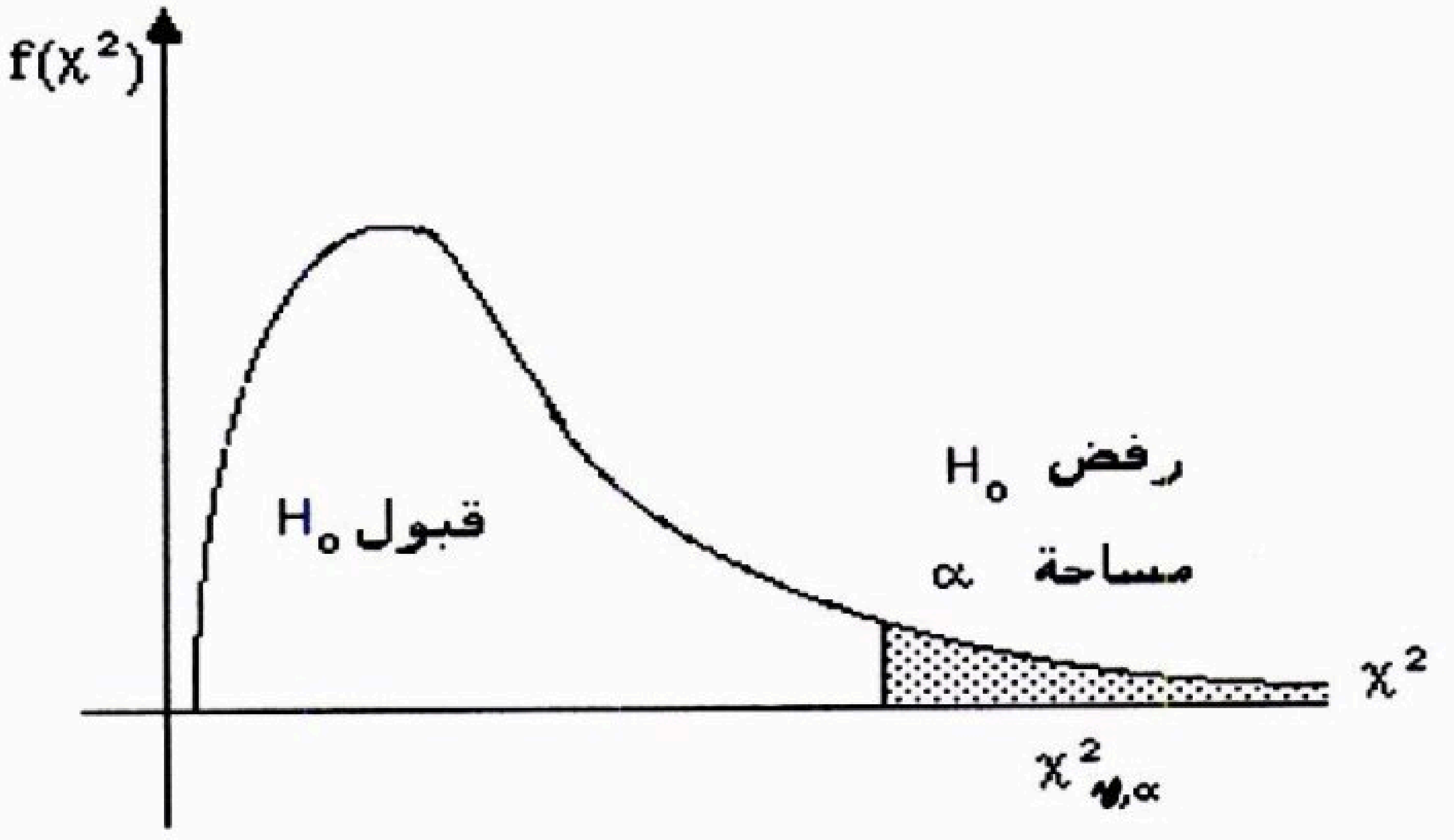
$$\sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \approx \chi^2_{\nu} \qquad (٨,١)$$

يكون لها توزيع يقترب من توزيع χ^2 بدرجات حرية $\upsilon=(k-1)$ حيث O_i ترمز إلى التكرارات المشاهدة للنتيجة رقم i و E_i ترمز إلى التكرارات المتوقعة المناظـرة للنتيجة رقم i وأن $n=\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i$ هي حجم العينة.

ويكـون هذا التقريـب مقبولا إذا كان عدد المشاهدات في العينة محـــل الدراسة $n \geq 50$، والتكرار المتوقع المناظر لكل نتيجة لا يقل عن خمسة مشاهدات وأحيانا عن مشاهدة واحدة على الأقل.

ملاحظة

إذا كان التكرار المشاهد يساوي التكرار المتوقع، وذلك لكل نتيجة من النتائج فإن قيمة الإحصــاءة χ^2 تكون مساوية للصفر ويمكن في هذه الحالة قبــول فـرض العدم H_0 ولكن إذا اختلف التكرار المـشـاهـد O_i والتكرار المتوقـع E_i فإن قيمـة الإحصاءة χ^2 تكون أكبر من الصفر وتزداد هذه القيمة كلمـا زاد الاخـتـلاف بـين التكرارات المشاهدة والمتوقعة. وفي هذه الحالة نرفض فرض العدم H_0 ونقبل الفرض البديل H_1. وحيث إننا نرفض فرض العدم H_0 عندما تكون قيمة الإحصاءة للاختبار كبيرة فإن منطقة الرفض تكون دائما في الطرف الأيمن من المنحنى الاحتمالي لتوزيع χ^2. أي أن الاختبار في هذه الحالة يكون اختبار طرف واحد هو الطرف الأيمن ونوضـح منطقة قبـول H_0 بالفتـرة $(0, \chi^2_{\upsilon,\alpha})$ حيث υ يسمى درجات الحريـة و α مستوى المعنوية وكذلك منطقة الرفض H_0 بالجزء المظلل في شكل (٨,١) كما يلي:



شكل (١ ,٨)

(١ ,٢ ,٨) اختبار حسن المطابقة لنظرية معينة

ونوضح ذلك من خلال المثالين التاليين

مثال (١ ,٨)

تم إلقاء قطعة عملة مائة مرة فحصلنا على النتائج التالية (58 صورة و 42 كتابة). فهل تتفق هذه المشاهدات مع كون العملة متزنة ؟ وذلك عند مستوى معنوية $\alpha = 0.05$.

الحل

أولا: صياغة الفرض الإحصائي.

العملة متزنة: H_0

العملة غيرمتزنة: H_1

$$\alpha = 0.05$$

ثانيا: اختبار الإحصاءة ثم حسابها.

نختار الإحصاءة $$\chi_0^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i}$$

حيث O_i هو التكرار المشاهد لإلقاء قطعة العملة.

E_i هو التكرار المتوقع لقطعة العملة على فرض أنها متزنة.

υ درجات الحرية وهي عدد النتائج مطروحا منها واحدا.

تحت صحة الفرض H_0 إذا كانت العملة متزنة فإننا نحصل 50% صورة وعلى 50% كتابة ونوضح التكرارات المشاهدة O_i والمتوقعة E_i بالجدول (١ ,٨) كما يلي:

جدول (١ ,٨)

التكرار \ النتائج	صورة	كتابة
O_i التكرار المشاهد	58	42
E_i التكرار المتوقع	50	50

من جدول (١ ,٨) يمكن حساب χ_0^2 كما يلي:

$$\chi_0^2 = \frac{(58-50)^2}{50} + \frac{(42-50)^2}{50} = 2.56$$

ودرجات الحرية (υ) وهي عدد النتائج وهي 2 في مثالنا (كتابة، صورة) مطروحا منها واحدا أي أن $\upsilon = 2 - 1 = 1$.

ثالثا: تحديد منطقة قبول ورفض H_0.

نحسب $\chi^2_{\upsilon,\alpha}$ من جدول مربع كاي (٣) آخر الكتاب حيث $\upsilon = 1$ و $\alpha = 0.05$

فنجـد أن $\chi^2_{1,0.05} = 3.841$. وأن فترة القبـول H_0 هي($\chi^2_{1,0.05}$,0) وهي (0, 3.841) وفترة رفض H_0 هي خارج فترة القبول السابقة.

رابعا: القرار الإحصائي.

حيث إن $\chi^2_0 = 2.56$ المحسوبة واقعة داخل فترة قبول H_0 فالقرار هنا هـو أننا لا نستطيع رفض H_0 أي أن العملة متزنة.

مثال (٢ ,٨)

طبقا لإحدى نظريات الوراثة في علم النبات، إن أحد أنواع الزهور يعطي في الفرع الواحد نسبة 1 : 1 : 2 أزهار صفراء وبيضاء وحمراء على الترتيب قام أحد الباحثين بأخذ عينة من مائة حبة من هذا النوع من الزهور وتم زراعتها وحصر عدد الزهور من كل لون في الفرع الواحد فكانت 30 من اللون الأصفر و 30 زهرة من اللون الأبيض و 40 زهرة من اللون الأحمر. فهل تتفق هذه المعلومات مع افتراضات النظرية السابقة ؟ وذلك عند مستوى معنوية $\alpha = 0.01$.

الحل

أولا: صياغة الفرض الإحصائي.

H_0 : النظرية صحيحة والمعلومات المشاهدة تؤديها

H_1 : النظرية غير صحيحة والمعلومات لا تبرر قبولها

ثانيا: اختيار الإحصاءة ثم حسابها

نحسب التكرارات المتوقعة طبقا للنظرية كما يلي:

عدد الزهور الصفراء $E_1 = 100 \times \frac{1}{4} = 25$ زهرة

عدد الزهور البيضاء $E_2 = 100 \times \frac{1}{4} = 25$ زهرة

عدد الزهور الحمراء $E_3 = 100 \times \frac{2}{4} = 50$ زهرة

يوضح جدول (٢، ٨) كل من التكرارات المشاهدة O_i والتكرارات المتوقعة E_i كما يلي:

جدول (٢، ٨)

التكرار \ النتائج		زهور صفراء	زهور بيضاء	زهور حمراء
التكرار المشاهد	Q_i	30	20	40
التكرار المتوقع	E_i	25	25	50

درجات الحرية υ هي عدد النتائج مطرحا منها واحدا أي أن $\upsilon = 3-1 = 2$
نحسب الإحصاءة χ_0^2 بالعلاقة التالية وباستخدام المعلومات في جدول (٢، ٨) السابق كما يلي:

$$\chi_0^2 = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(30 - 25)^2}{25} + \frac{(30 - 25)^2}{25} + \frac{(40 - 50)^2}{50}$$

$$= 1 + 1 + 2 = 4$$

ثالثا: تحديد منطقة قبول ورفض H_0.

نختار مستوى المعنوية $\alpha = 0.01$ فإن $\chi^2_{2,0.01}$ من جدول مربع كاي (٣) آخر الكتاب هي $\chi^2_{2,0.01} = 9.21$ ويمكن تحديد فترة قبول H_0 هي $(0, \chi^2_{2,0.01})$ وهي $(0, 9.21)$ ومنطقة رفض H_0 خارج الفترة السابقة.

رابعا: القرار الإحصائي.

حيث إن قيمة $\chi_0^2 = 4$ المحسوبة واقعة داخل فترة قبول H_0 أي أننا لا نستطيع رفض H_0 أي أن النظرية صحيحة.

(٨,٢,٢) اختبار جودة التوفيق لتوزيع ذي الحدين

ولتوضيح اختيار جودة التوفيق لتوزيع ذي الحدين ندرس المثال التالي

مثال (٨,٣)

في دراسة لمراقبة جودة الإنتاج للآلات المستخدمة في رصف الطرق أخذ مهندس عينة يوميا مكونة من 4 آلات ولفترة 200 يوما وذلك من خط الإنتاج لهذه الآلات بالمصنع حيث يتم فحصها يوميا ويتم تسجيل عدد الآلات التي تحتاج إلى إعادة ضبط مرة أخرى خلال فترة 200 يوم السابقةوكانت البيانات في جدول (٨,٣) كما يلي:

جدول (٨,٣)

عدد الآلات التي أعيد ضبطها	0	1	2	3	4
عدد الأيام (التكرار O_i)	102	78	19	1	0

هل هذه البيانات تخضع لتوزيع ذي الحدين $X : b(4, 0.1)$ ؟

وحيث X يمثل عدد الآلات المطلوب إعادة ضبطها في العينة اليومية ($n = 4$).

وأن نسبة نجاح المطلوب إعادة ضبطه في العينة هي $P = 0.1$ وذلك عند مستوى معنوية $\alpha = 0.05$.

الحل

إذا كان نموذج ذي الحدين $X : b(4, 0.1)$ صحيحا فإنه يمكن حساب الاحتمالات

التاليةلاستخدامها في التكرار المتوقع (E_i) وذلك كما يلي:

$$P(x=0) = \binom{4}{0}(0.1)^0(0.9)^4 = 0.6561$$

$$E_1 = 200\,P(x=0) = 200\,(0.6561) = 131.22$$ يوم

وأن

$$P(x=1) = \binom{4}{1}(0.1)^1(0.9)^3 = 0.2916$$

$$E_2 = 200\,P(x=1) = 200\,(0.2916) = 58.32$$ يوم

وأن

$$P(x=2) = \binom{4}{2}(0.1)^2(0.9)^2 = 0.0486$$

$$E_3 = 200\,P(x=2) = 200\,(0.049) = 9.72$$ يوم

وأن

$$P(x=3) = \binom{4}{3}(0.1)^3(0.9)^1 = 0.0036$$

$$E_4 = 200\,P(x=3) = 200\,(0.004) = 0.72$$ يوم

وأن

$$P(x=4) = \binom{4}{4}(0.1)^4(0.9)^0 = 0.0001$$

$$E_5 = 200\,P(x=4) = 200\,(0.0001) = 0.02$$ يوم

ويمكن تلخيص القيم المشاهدة Q_i في جدول (٨,٣) السابق مع القيم المتوقعة E_i التي تم حسابها من توزيع ذي الحدين $b(4, 0.1)$ في جدول (٨,٤) كما يلي:

جدول (٨,٤)

عدد الآلات التي أعيد ضبطها	0	1	2	3	4
عدد الأيام المشاهدة(التكرار O_i)	102	78	19	1	0
عدد الأيام المتوقعة (التكرار E_i)	131.22	58.32	9.72	0.72	0.02

ملاحظة

في اختبارات جودة التوفيق (حسن المطابقة) إذا كانت أحد الفئات لها تكرار متوقع أقل من الواحد فإن هذه الفئة تدمج مع الفئة المجاورة لها لكي يكون الناتج تكرار أكبر من الواحد الصحيح، وعلى ذلك يمكن دمج الفئات التي فيها عدد الآلات 2 ، 3 و 4 والتي تم إعادة ضبطها في فئة واحدة وبذلك يصبح جدول (٨,٤) السابق كما هو موضح في الجدول (٨,٥) كما يلي:

جدول (٨,٥)

عدد الآلات التي أعيد ضبتها	0	1	≥2
عدد الأيام المشاهدة (التكرار Q_i)	102	78	20
عدد الأيام المتوقعة (التكرار E_i)	131.22	58.32	10.46

ونستخدم جدول (٨,٥) السابق لحساب الاحصاءة

$$\chi_0^2 = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i}$$

واختبارها في الخطوات التالية:

أولا: صياغة الفرض الإحصائي.

البيانات تتبع توزيع ذي الحدين $H_0 : b(4, 0.1)$

البيانات لا تتبع توزيع ذي الحدين $H_1 : b(4, 0.1)$

$\alpha = 0.05$

ثانيا: اختيار الإحصاءة ثم حسابها.

نختار الإحصاءة $\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$

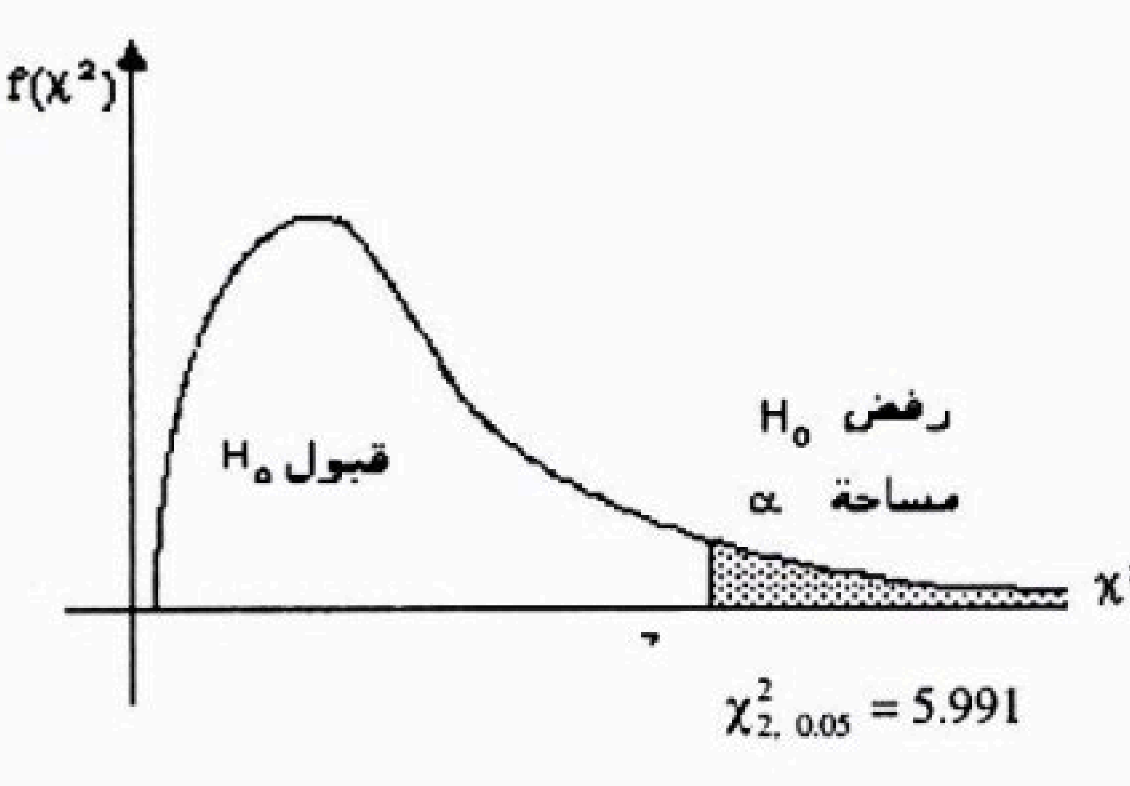
ويتم حسـابها من جدول (٨,٥) السابق كما يلي:

$$\chi_8^2 = \frac{(102-131.22)^2}{131.22} + \frac{(78-58.32)^2}{58.32} + \frac{(20-10.46)^2}{10.48} = 21.41$$

ثالثا: تحديد منطقة قبول ورفض H_0.

يتم تظليل منطقة رفض H_0 كما هو موضح في شكل (٨,٢) التالي:

من شكل (٨,٢) نجد أن فترة قبول H_0 هي $(0, \chi^2_{\upsilon,\alpha})$ حيث درجات الحرية $(\upsilon =$ عدد الفئات - عدد المعالم المقدرة - $1)$ $\upsilon = K - 1 = 3 - 1 = 2$



شكل (٨,٢)

فتكون فترة قبول H_0 هي $(0, \chi^2_{2, 0.05})$ ومن جدول توزيع χ^2 رقم (٣) في آخر الكتاب نجد أن الفترة هي $(0, 5.991)$ ويكون رفض H_0 خارج هذه الفترة.

رابعا: القرار الإحصـائي :

نظرا لأن قيمة كاي المحسوبة هي 21.41 واقعة خارج فترة القبول H_0 والتي هي(0, 5.991) .

إذاً القرار هو رفض H_0 أي أن البيانات لا تتبع توزيع ذي الحدين b(4 , 0.1)

(٣ , ٢ , ٨) اختبار جودة التوفيق لتوزيع بواسون

ولتوضيح اختبار جودة التوفيق(حسن المطابقة) لتوزيع بواسون نـوضـح ذلك من خلال المثال التالي.

مثال (٤ , ٨)

في دراسة لمعرفة عدد الأخطاء في الدوائر الكهربائية لأحرف لوحة طابعة اختيرت عينة مكونة من عدد $n = 60$ لوحة طابعـة وعدد العيـوب التي تم الحصـول عليها سجـل في جدول (٦ , ٨) والمطلوب معرفة ما إذا كان عدد الأخطاء في هذه الأحرف يتبع توزيع بواسون أم لا وذلك عند مستوى معنوية $\alpha = 0.05$.

جدول (٦ , ٨)

عدد عيوب أحرف الطباعة	0	1	2	3
عدد االلوحات التي بها عيوب (التكرار O_i)	32	15	9	4

الحل

يحسب متوسط توزيع بواسون من بيانات الجدول (٦ , ٨) السابق كما يلي :

$$\lambda = \overline{x} = \left[32(0) + 15(1) + 9(2) + 3(4)\right] / 60 = 0.75$$

ولحساب التكرار المتوقـع E_i نستخدم الاحتمالات من توزيع بواسون كمـا يلي :

$$P(x=k) = \frac{(0.75)^{k} e^{-0.75}}{k!} \quad , \quad k = 0\ , 1, 2, 3$$

وتكون القيمة المتوقعة E_i هي:

$$E_1 = 60\left[P(x=0)\right] = 60\left[\frac{(0.75)^{0} e^{-0.75}}{0!}\right]$$

$$= 60\ (0.472) = 28.32$$

$$E_2 = 60\left[P(x=1)\right] = 60\left[\frac{(0.75)^{1} e^{-0.75}}{1!}\right]$$

$$= 60\ (0.354) = 21.24$$

$$E_3 = 60\left[P(x=2)\right] = 60\left[\frac{(0.75)^{2} e^{-0.75}}{2!}\right]$$

$$= 60\ (0.133) = 7.98$$

$$E_4 = 60\left[P(x=3)\right] = 60\left[\frac{(0.75)^{3} e^{-0.75}}{3!}\right]$$

$$= 60\ (0.041) = 2.46$$

ويمكن تلخيص القيم المشاهدة O_i في الجدول (٨,٦) والقيم المتوقعة E_i التي تم حسابها في الجدول (٨,٧) كما يلي:

جدول (٨,٧)

عدد عيوب أحرف الطباعة	0	1	2	3
عدد الطابعات التي بها عيوب (التكرار O_i)	32	15	9	4
عدد الطابعات المتوقع بها عيوب (التكرار E_i)	28.24	21.24	7.98	2.49

ويستخدم جدول (٨,٧) السابق لحساب الإحصاءة $\chi_0^2 = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i}$

واختبارها كما هو موضح في الخطوات التـالية:

أولا: صياغة الفرض الإحصائي

H_0 : توزيع عيوب الأخطاء لأحرف الطباعة تتبع توزيع بواسـوان

H_1 : توزيع عيوب الأخطاء لأحرف الطباعةلا تتبع توزيع بواسوان

$\alpha = 0.05$

ثانيا: اختيار الإحصاءة ثم حسابها

من جدول (٨,٧) يمكن حساب الإحصاءة χ_0^2 كما يلي:

$$\chi_0^2 = \sum_{i=1}^{8} \frac{(O_i - E_i)^2}{E_i} = \frac{(32-28.32)^2}{28.32} + \frac{(15-21.24)^2}{21.24} + \frac{(9-7.98)^2}{7.98} + \frac{(4-2.49)^2}{2.49}$$

$$= 3.357$$

ثالثا: تحديد منطقة قبول ورفض H_0

يتم تظليل منطقة رفض$_0$ H كما هو موضح في شكل (٨,٣) التالي:

من شكل (٨,٣) نجد أن فترة قبول$_0$ H هي$(0.\chi^2_{\upsilon,\alpha})$

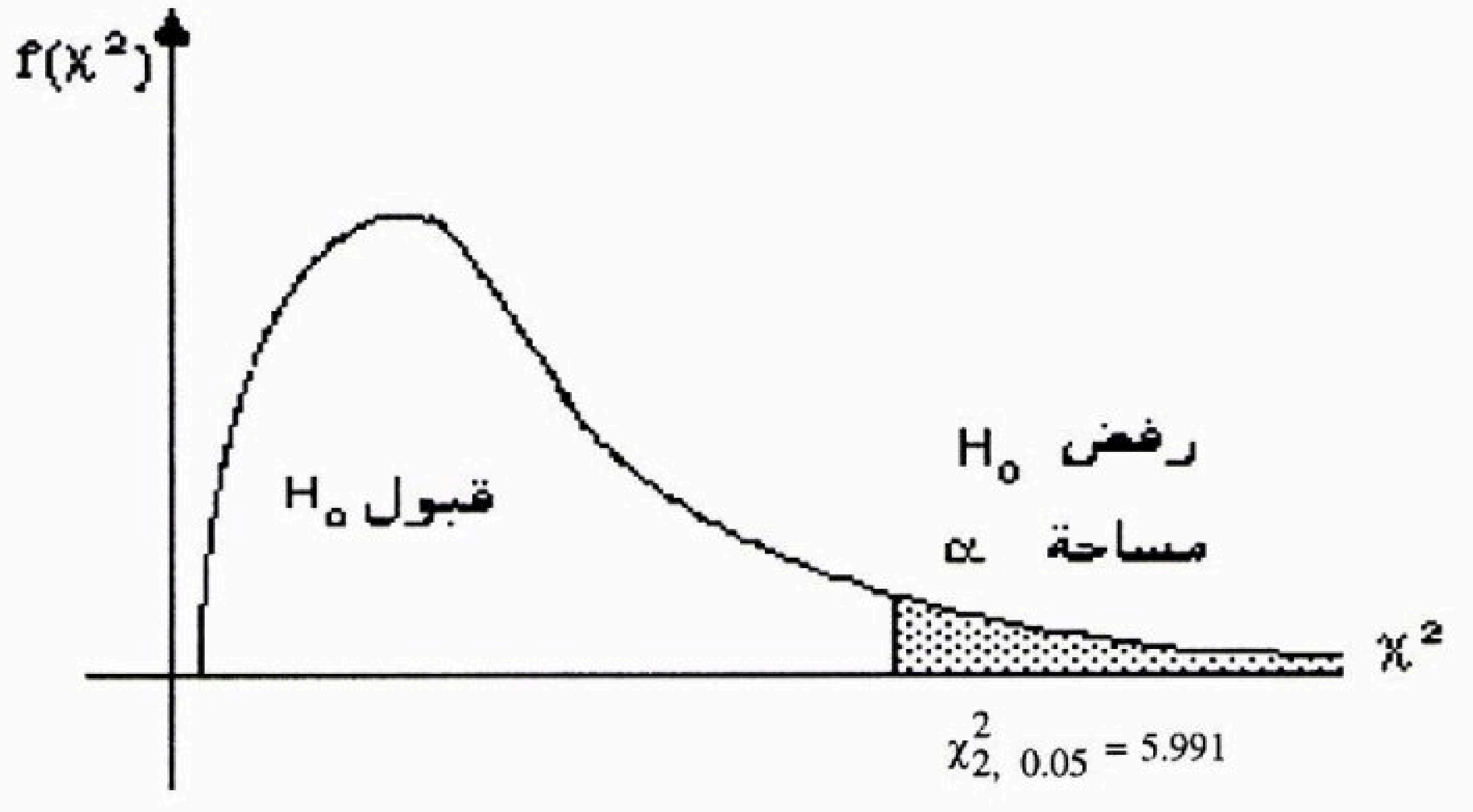
حيث درجات الحرية $2 = 1 - 1 - 4$ = (1 - عدد المعالم المقدرة - عدد الفئات = υ)

فتكون فترة قبــول$_0$ H هي$(0, \chi^2_{2\,,\,0.05})$ ومن جدول توزيع χ^2 رقم (٣) في آخر الكتاب نجد أن الفترة هي$(0, 5.991)$ ويكون رفض$_0$ H خارج هذه الفترة.

رابعا: القرار الإحصائي.

نلاحظ قيمة مربع كاي المحسوبة $\chi_2^2 = 3.357$ واقعة داخل فترة قبول$_0$ H

وهي $(0, \chi^2_{2\,,\,0.05})$ومن ذلك فإننا لا نستطيع رفض H_0، أي أن توزيع الأخطاءلأحرف لوحات الطابعات تتبع توزيع بواسون بمعلمة $\lambda = 0.75$.



شكل (٨,٣)

(٨,٢,٤) اختبار جودة التوفيق للتوزيع الطبيعي

ولاختبار جودة التوفيق (حسن المطابقة) للتوزيع الطبيعي نوضح ذلك من خلال المثال التالي.

مثال (٨,٥)

أخذت عينة من سائل المخ لعدد مكون من 100 مريض يشكون من اضطراب معين. وتم تحليل هذه العينة من هذه السوائل وتسجيل درجة التركيز للأحماض الأمينية للمرضى وتم تلخيص النتائج في الجدول التكراري (٨,٨) التالي:

جدول (٨ , ٨)

فئات تركيز الأحماض للمرضى	أقل من 4.4	–4.4	–4.84	–5.24	≥ 5.68	المجموع
عدد المرضى (التكرار المشاهد O_i)	22	17	14	24	23	100

والمطلوب معرفة ما إذا كان درجة تركيز الأحماض الأمينية في المخ لهؤلاء المرضى تتبع التوزيع الطبيعي $N(5.04 , 0.64)$ وذلك عند مستوى معنوية $\alpha = 0.01$. حيث تم تقدير كل من μ, σ^2 من العينة.

الحل

من المثال (٥ , ٨) نجد أن المتوسط $\hat{\mu} = 5.04$ وأن التباين $S^2 = 0.64$ وأن الانحراف المعياري $S = 0.8$. وعدد المرضى O_i لكل خلية والمبين في الجدول (٨ , ٨) السابق.

والمطلوب الآن حساب عدد المرضى المتوقع E_i حتى يمكن استخدام اختبار χ^2 التي تعطي قيمة χ_0^2 تحت صحة فرض العدم H_0 بالصيغة التالية:

$$\chi_0^2 = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i}$$

وتحسب القيمة المتوقعة E_1 باستخدام معلومات الجدول (٨ , ٨) السابق وجدول التوزيع الطبيعي رقم (١) آخر الكتاب كما يلي:

$$\begin{aligned} E_1 &= nP(x \leq x_1) \\ &= 100P(-\infty < x < 4.4) \\ &= 100P\left(-\infty < z < \frac{4.4 - 5.04}{0.8}\right) \\ &= 100P(-\infty < z < -0.8) \\ &= 100 \quad (0.2119) = 21.19 \end{aligned}$$ مريض

وتحسب القيمة المتوقعة E_2 بنفس الطريقة كما يلي:

$$E_2 = n\,P(-4.4 < x \le 4.84)$$
$$= 100\ P(-0.8 < z < 0.25)$$
$$= 100\ (0.1894) = 18.94$$ مريض

وتحسب القيمة المتوقعة E_3 بنفس الطريقة كما يلي:

$$E_3 = n\,P(4.84 < x \le 5.24)$$
$$= 100\ P(-0.25 < z < 0.25)$$
$$= 100\ (0.1974) = 19.74$$ مريض

وتحسب القيمة المتوقعة E_4 بنفس الطريقة كما يلي:

$$E_4 = n\,P(5.24 < x \le 5.68)$$
$$= 100\ P(0.25 < z < 0.8)$$
$$= 100\ (0.1894) = 18.94$$ مريض

وتحسب القيمة المتوقعة E_5 بنفس الطريقة كما يلي:

$$E_5 = nP(5.68 < x \le \infty)$$
$$= 100\,P(0.8 < z < \infty)$$
$$= 100\ (0.2119) = 21.19$$ مريض

ويمكن تلخيص القيم المشاهدة في جدول (٨,٨) السابق والقيم المتوقعة التي تم حسابها في جدول (٨,٩) كما يلي:

جدول (٨,٩)

فئات تركيز الأحماض للمرضى	أقل من 4.4	4.4 –	4.84 –	5.24 –	≥ 5.68	المجموع
عدد المرضى (التكرار المشاهد O_i)	22	17	14	24	23	100
عدد المرضى (التكرار المتوقع E_i)	21.19	18.94	19.74	18.94	21.19	100

ونستخدم جدول (٨,٩) السـابق لحسـاب الإحصـاءة $\chi_0^2 = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i}$

واختبارها كما هو موضح في الخطوات التالية:

أولا: صياغة الفرض الإحصائي.

H_0: توزيع الأحماض الأمينية للمرضى يتبع توزيع طبيعي

H_1: توزيع الأحماض الأمينية للمرضى لا يتبع توزيع طبيعي

$\alpha = 0.01$

ثانيا: اختيار الإحصاءة ثم حسابها.

نختار الإحصـاءة χ_0^2 تحت صحة فرض العـدم H_0، وذلك باستخـدام الجدول (٨,٩) السابق كما يلي:

$$\chi_0^2 = \frac{(22-21.19)^2}{21.19} + \frac{(17-18.94)^2}{18.94} + \frac{(14-19.74)^2}{19.74} + \frac{(24-18.94)^2}{18.94} + \frac{(23-21.19)^2}{21.19}$$

$$=0.031+0.199+1.669+1.901+0.155= 3.955$$

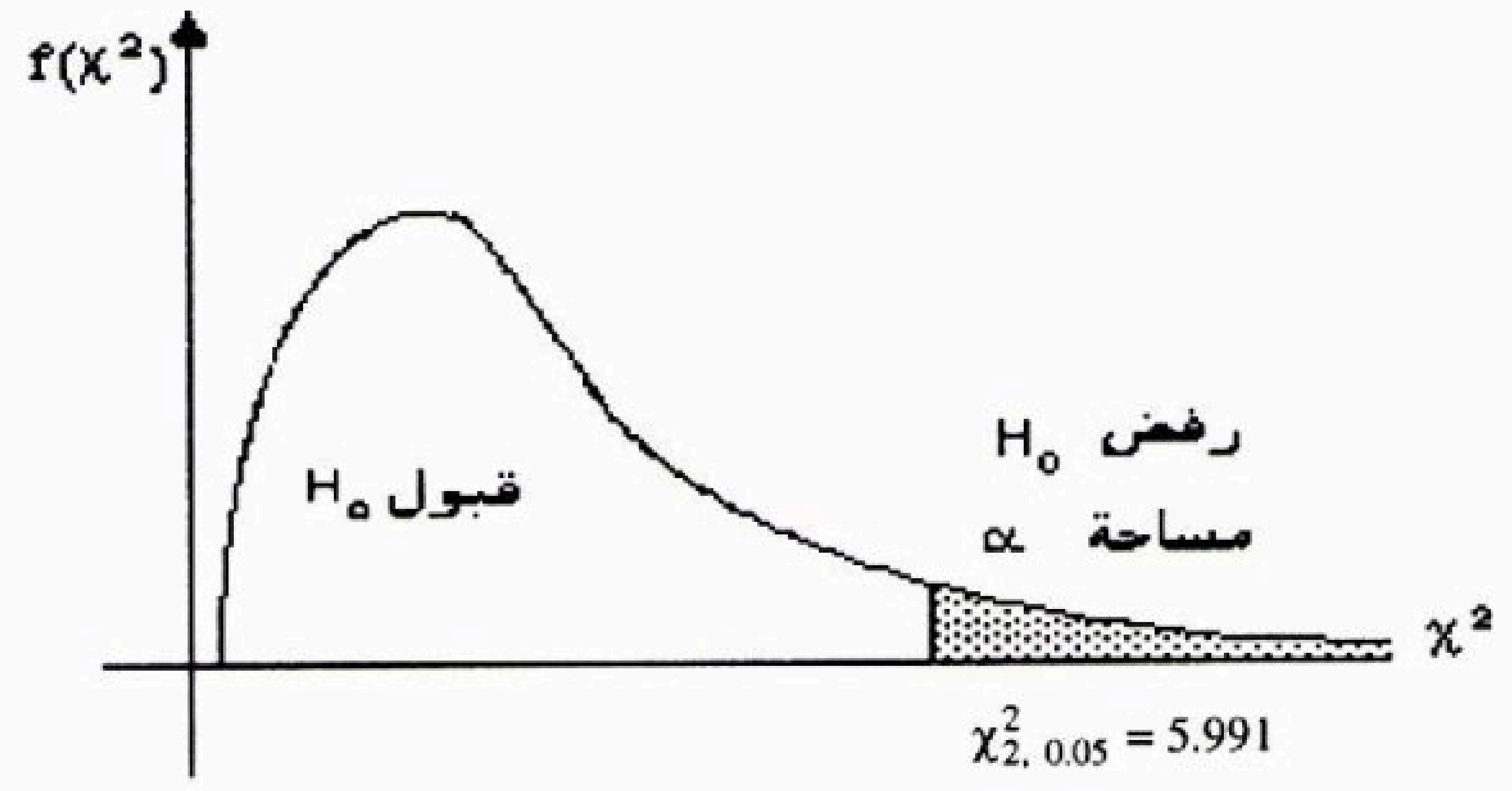
وتحسب درجـات الحريـة من العلاقـة $\upsilon = k - m - 1$ حيث $k = 5$ عـدد الفـئات و $m = 2$ لأننا قدرنا المتوسط μ والانحراف المعياري σ وتصبح $\upsilon = 5 - 2 - 1 = 2$.

ثالثا: تحديد منطقة قبول ورفض H_0.

يتم تظليل منطقة رفض H_0 كما هو موضح في شكل (٨,٤) التالي:

من شكل (٨,٤) نجد أن فترة قبول H_0 هي $\left(0, \chi^2_{\upsilon,\alpha}\right)$ حيث درجات الحرية

$(\upsilon =$ عدد الفئات $-$ عدد المعالم المقدرة $- 1) = 5 - 2 - 1 = 2$



شكل (٤ ,٨)

فتكون فترة قبـول H_0 هي$(0, \chi^2_{2\,,\,0.01})$ ومن جدول توزيع χ^2 رقم (٣) في آخر الكتاب نجد أن الفترة هي$(0, 9.210)$ ويكون رفض H_0 خارج هذه الفترة.

رابعا: القرار الإحصائي.

نلاحظ قيمة مربع كاي المحسوبة $\chi^2_0 = 3.955$ واقعة داخل فترة قبول H_0 وهي $(0, \chi^2_{2\,,\,0.01})$وهي $(0, 9.210)$ومن ذلك فإننا لا نستطيع رفض H_0، أي أن توزيع تركيز الأحماض الأمينية لمرضي العينة في السائل حول المخ يتبع توزيع الطبيعي $N(5.04, 0.64)$.

ملاحظة

توجد طريقة أخرى لاختبار حسن المطابقة للتوزيع الطبيعي ويفضل استخدام هذه الطريقة مع التوزيعات المتصلة (مثل الطبيعي والأسي. . الخ). وهذه الطريقة

تبين قوة التفريق بين التكرار المشاهد O_i والتكرار المتوقع E_i لاختبار مربع كاي.
يختار في هذه الطريقة فئات التوزيع المتصل تحت صحة فرض العدم H_0 بحيث يكون لها احتمالات متساوية، أي أن التكرارات المتوقعة E_i في جميع الفئات متساوية ونوضح طريقة حساب واستخدام هذه الطريقة من خلال المثال التالي:

مثال (٦ ,٨)

البيانات التالية هي الإنتاج اليومي من اللبن بالكجم لعدد 40 بقرة (من نفس العمر والجنس) ويعتقد أن هذه البيانات تتبع توزيع طبيعي

16.93	18.79	14.62	13.98	15.79	12.39	13.20	16.08	13.97	16.16
16.12	17.81	18.74	15.99	13.32	13.43	16.40	13.76	16.58	15.25
18.97	18.36	15.04	18.79	18.08	17.32	16.32	17.54	18.05	14.20
18.04	13.00	13.25	12.43	16.56	14.12	20.55	16.75	13.29	18.23

والمطلوب اختبار ما إذا كانت هذه المشاهدات تتبع توزيع طبيعي عند مستوى معنوية $\alpha = 0.05$.

الحل

أولا: صياغة الفرض الإحصائي.

H_0: البيانات تتبع توريع طبيعي

H_1: البيانات لا تتبع توزيع طبيعي

ثانيا: اختيار الإحصاءة وحسابها.

نختار الإحصاءة $\chi_0^2 = \dfrac{\sum_{i=1}^{k} (O_i - E_i)^2}{E_i}$ ويتم حسابها بحساب التكرارات

المشاهدة O_i لكل فئة وكذلك التكرارات المتوقعة E_i لكل فئة وفي هذه الطريقة حيث إن عدد البيانات $n = 40$ فإننا نختار 8 فئات وبما أن التكرارات المتوقعة في جميع الفئات متساوية فإن $E_i = \frac{40}{8} = 5 \ , \ i = 1,2,...,8$ (لاحظ أن 5 هو أقل تكرار متوقع يسمح به في أية فئة أحيانا). بما أن احتمال كل فئة هو نفسه فإن هذا الاحتمال يجب أن يساوي 1/8 أي 0.125 وهكذا فإن الحد الأعلى للفئة الأولى هو المئين ذو الرتبه 12.5 والحد الأعلى للفئة الثانية هو المئين ذو الرتبة $25 = 12.5 + 12.5$ والحد الأعلى للفئة الثالثة هو المئين ذو الرتبة 37.5 وهكذا حتى لآخر حد الفئة الثامنة ونوضح الحسابات لتعيين المئينات المطلوبة للتوزيع الطبيعي فيلزم حساب المتوسط $\overline{x}$ والانحراف المعياري s من البيانات السابقة حيث إن قيمتهما $s = 2.144$ و $\overline{x} = 15.9$ ونحسب حدود الفئات باستخراج قيمة المئين $Z(a_i)$, $i=1,2,...,7$ من جدول التوزيع الطبيعي (١) آخر الكتاب فنحصل على حدود الفئات $x_i \ , \ i = 1,2,..., 7$ حيث إن:

$$x_i = \overline{x} + z(a_i)s$$

وبذلك فإن قيمة$z(a_1)$ تمثل المئيني الذي على يساره مساحة قدرها 12.5%وتكتب

$$z(a_1) = z(0.125)$$

وكذلك فإن قيمة$z(a_2)$ تمثل المئيني الذي على يساره مساحة قدرها 25 % وتكتب

$$z(a_2) = z(0.25)$$

وهكذا لباقي القيم كما هو موضح فيما يلي:

$$x_1 = \overline{x} + Z(0.125)s = 15.96 - 1.15(2.144) = 13.49$$

$$x_2 = \overline{x} + Z(0.250)s = 15.96 - 0.67(2.144) = 14.52$$

$$x_3 = \overline{x} + Z(0.375)s = 15.96 - 0.32(2.144) = 15.27$$

$$x_4 = \overline{x} + Z(0.500)s = 15.96 - 0.00(2.144) = 15.96$$

$$x_5 = \bar{x} + Z(0.625)s = 15.96 + 0.32(2.144) = 16.65$$

$$x_6 = \bar{x} + Z(0.750)s = 15.96 + 0.67(2.144) = 17.40$$

$$x_7 = \bar{x} + Z(0.857)s = 15.96 + 1.15(2.144) = 18.43$$

وبعد حساب حدود الفئات x_i يمكن صياغة جدول (٨,١٠) بحيث يشتمل على حدود الفئات المحسوبة أعلاه وكذلك التكرار المشاهد O_i من بيانات في المثال (٨,٦) السابق والتكرار المتوقع المتساو وهو $E_i = 5$ وكذلك حساب القيمة $\frac{(O_i - E_i)^2}{E_i}$ لكل فئة كما يلي:

جدول (٨,١٠)

حدود الفئات لإنتاج الحليب	< 13.49	13.49 – 14.51	14.52 – 15.26	15.27 – 15.95	15.96 – 16.64	16.65 – 17.39	17.40 – 18.42	≥ 18.43
التكرار المشاهد O_i	7	6	3	1	8	3	7	5
التكرار المتوقع E_i	5	5	5	5	5	5	5	5
$\frac{(O_i - E_i)^2}{E_i}$	0.80	0.20	0.80	3.20	1.80	0.80	0.80	0.00

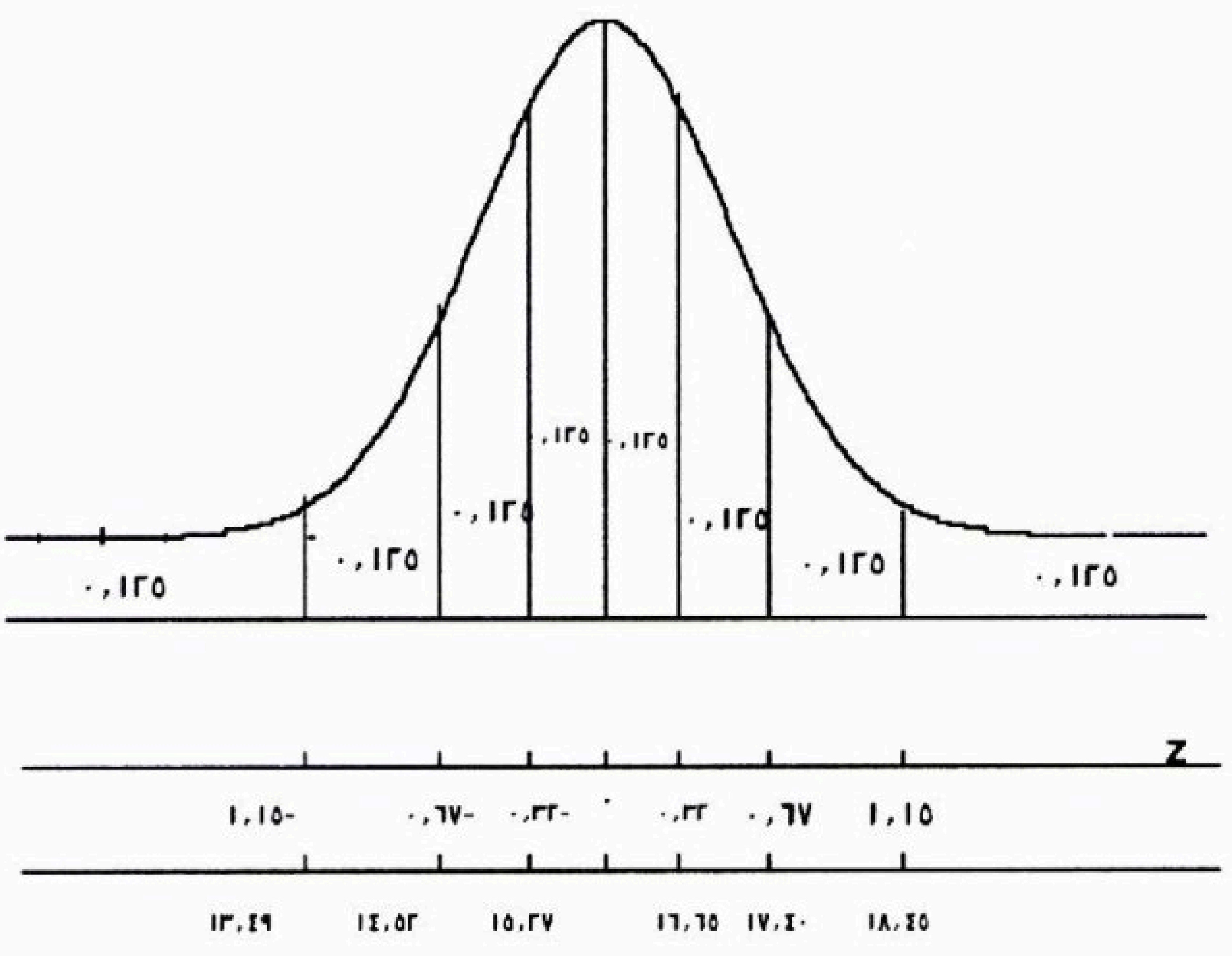
من جدول (٨,١٠) نجد أن قيمة الإحصاءة χ_0^2 هي

$$\chi_0^2 = \sum_{i=1}^{8} \frac{(O_i - E_i)^2}{E_i} = 8.4$$

حيث يتم الحساب للتكرار المشاهد O_i من البيانات المشاهدة في المثال (٨,٦) السابق وعلى سبيل المثال التكرار المشاهد O_1= عدد المشاهدات التي تقل عن 13.49 يساوي

7 مشاهدات وهكذا لباقي الفئات في جدول (٨,١٠) ولحساب درجات الحرية υ من العلاقة $\upsilon = k - m - 1$ حيث $k = 8$ عدد الفئـات و $m = 2$ حيث تم تقدير المتوسط μ والانحراف المعياري σ وبذلك فإن $\upsilon = 8 - 2 - 1 = 5$.

وتوضح قيم المئينات x_i السابقه بالرسم حسب شكل (٨,٥) كما يلي:



شكل (٨,٥)

ثالثا: تحديد منطقة قبول ورفض H_0

تحدد فترة قبول H_0 بالفترة $(0, \chi^2_{5,\ 0.05})$ حيث يتم استخراج $\chi^2_{2,\ 0.05} = 11.071$ من جدول مربع كاي رقم (٣) آخر الكتاب وبذلك تكون فترة قبول H_0 هي $(0, 11.071)$

ومنطقة رفض H_0 خارج هذه الفترة.

رابعا: القرار الإحصائي.

حيث قيمة مربع كاي المحسوبة $\chi_0^2 = 8.4$ واقعة داخل فترة قبول H_0 وهي $(0, 11.071)$ وعليه فإننا لا نستطيع رفض H_0، أي أن البيانات المشاهدة تتبع التوزيع الطبيعي بمتوسط $\mu = 15.96$ وانحراف معياري $\sigma = 2.144$.

(٨,٣) اختبارات مربع كاي للاستقلال والتجانس

Chi-square Testing of Independence and Homogineity

يعتبر اختبارات مربع كاي للأستقلال والتجانس في جداول التوافق من أهم وأكثر استخدامات مربع كاي لاختبار صحة فرض العدم H_0 بأن معياريي التصنيف عند تطبيقهما على نفس الوحدات في مجتمع ما يكونا مستقلين (على سبيل المثال أخذت عينة من مجتمع ما لدراسة استقلال التدخين عن التعليم فإنه يتم تصنيف العينة وفق المعيار الأول بأن الشخص يدخن أو لا يدخن وحسب المعيار الثاني أن الشخص متعلم أو غير متعلم). ونقول إن لدينا معيارين للتصنيف مستقلين إذا كان توزيع أحد المعيارين هو نفسه مهما كان توزيع المعيار الآخر. وبذلك فإن فرض العدم H_0 للاستقلال هو أن معياري التصنيف مستقلين، ضد الفرض البديل H_1 أنهما غير مستقلين. وأما فرض العدم H_0 للتجانس هو أن العينات المسحوبة من مجتمعات متجانسة بالنسبة لمعياري التصنيف (أي أنه لا يوجد فرق بين الطبقات المختلفة المصنفة حسب هذين المعيارين). وفيما يلي نوضح كيفية تمييز كل من اختبارات الاستقلال والتجانس.

(٨,٣,١) اختبارات الاستقلال.

● نختار عينة واحدة من المجتمع ومن ثم تصنف وحدات العينة على أساس معياري التصنيف (أي أن المعاينة تسبق التصنيف).

● مجاميع الصفوف الأفقية أو الأعمدة الرأسية كما هو موضح في جدول التوافق (٨,١١) التالي تخضع للصدفة ولا تثبت مسبقا. أي أن عدد الوحدات المشاهدة الواقعة في أي خلية في جدول التوافق (٨,١١) تعين بعد سحب العينة وعليه فإن مجموع الصفوف أو الأعمدة هي كميات خاضعة للصدفة وليست تحت سيطرة الباحث.

(٨,٣,٢) اختبارات التجانس

● يختار مجتمعان أو أكثر وتسحب من كل منهما عينة مستقلة (أي أن التصنيف يسبق المعاينة).

● مجاميع الصفوف الأفقية (أو الأعمدة الرأسية) في جدول التوافق (٨,١١) يمكن أن تثبت مسبقا من قبل الباحث. وبذلك يمكن للباحث أن يعين مسبقا العينات المستقلة والتي تسحب من كل مجتمع مختار وعليه فإنه يمكن تثبيت واحد من المجاميع للمعيارين سواء كانت الصفوف أو الأعمدة مسبقا بينما يبقى الآخر عشوائيا.

● بعد إجراء المعاينة تصبح كل مشاهدة مصنفة حسب معياري التصنيف المعطاة.

(٨,٣,٣) تحليل جداول التوافق

لتحليل جداول التوافق (analysis of r x c association tables) لنفترض أن أي مشاهدة (أو وحدة تجربة) في مجتمع إحصائي S يمكن تصنيفه بواسطة معيارين مختلفين A, B. سوف نفترض أن المعيار A له $r \geq 2$ من المستويات أو الصفات والمعيار B له $c \geq 2$ من المستويات أو الصفات. فإذا وضعنا r صفة للمعيار A كصفوف و c صفة للمعيار B كأعمدة فإننا نكون جدول يسمى جدول التوافق $r \times c$ كما هو موضح في جدول (٨,١١) كما يلي:

جدول (٨,١١)

معيار التصنيف الأول A	معيار التصنيف الثاني B					مجموع الصفوف
	1	2	3	...	c	
1	O_{11}	O_{12}	O_{13}		O_{1c}	$O_{1.}$
2	O_{21}	O_{22}	O_{23}		O_{2c}	$O_{2.}$
:	:	:	:		:	
:	:	:	:		:	
r	O_{r1}	O_{r2}	O_{r3}		O_{rc}	$O_{r.}$
مجموع الأعمدة	$O_{.1}$	$O_{.2}$	$O_{.3}$		$O_{.c}$	$O_{..}=n$

ونلاحظ في جدول التوافق (٨,١١) أن تقاطع صف رقم i مع العمود رقم j يسمى خلية ويكون عدد المشاهدات لهذه الخلية هو O_{ij} والتي هي من المستوى i في المعيار A والمستوى j في المعيار B توضح في الخلية (i,j) حيث $j = 1,2,...,c$ و $i = 1, 2,..., r$ وتسمى O_{ij} التكرار المشاهد للخلية (i, j) وأن مجموع جميع الخلايا في الصف i يرمز له $O_{i.}$ ومجموع الخلايا في العمود j يرمز له $O_{.j}$ أي

أن $O_{i.} = \sum_{j=1}^{c} O_{ij}$, $i = 1, 2,..., r$ وأن $O_{.j} = \sum_{i=1}^{r} O_{ij}$, $j = 1, 2 ,..., c$

وأن مجموع مجاميع الصفوف والأعمدة يعطي المجموع الكلي أو n (حجم العينة) أي أن

$$n = O_{..} = \sum_{i=1}^{c} O_{i.} = \sum_{j=1}^{r} O_{.j}$$

بعد إنشاء جدول التوافق $r \times c$ فإنه يمكن اختبار الفروض حول المعيارين A, B ومنها مثلا استقلال المعيار A عن المعيار B، أو إذا كان المعيار A يقسم المجتمع S إلى مجموعات جزئية منفصلة $S_1, S_2 ,..., S_r$ فإن فرض الاستقلال يصبح فرض التجانس كما ذكرنا سابقا، فإن مستويات المعيار A تقسم المجتمع S إلى r

صفة تكّون المجموعات الجزئية $S_1 , S_2 ,..., S_r$. أي أن فرد أو مشاهدة من S_i هذه تصنف أو تتبع صفة واحدة أو مستوى واحد فقط من المستويات C للمعيار B. لنرمز p_{ij} لنسبة الأفراد في المجموعة S_i والتي تصنف في المستوى j للمعيار B. وبذلك فإن فرض العدم H_0 للتجانس هو

$$H_0 : p_{1j} = p_{2j} = ... = p_{rj} \; , \; j = 1,2 ,..., c$$

أي أن لأي مستوى j فإن نسبة الأفراد في مجموعة جزئيـة S_i هي نفسهـا وأن الفرض البديل H_1 يعطي نسبة واحدة على الأقل ليست مساوية لبقية النسـب. فرض الاستقلال في جدول (٨ , ١١) حيث لدينا مجتمع واحد S وكل فرد في هذا المجتمع يصنف بطريقتين حسب المعـيـاريـن A, B. في هذه الحالة يكون فـرض العدم H_0 للاستقلال هو

H_0 : المعيارين A, B مستقلين

ضد الفرض البديل H_1 وهو

H_1 : المعيارين A, B غير مستقلين

(٨ , ٣ , ٤) اختبار مربع كاي لجداول التوافق

لاختبار أي من فروض التجانس أو الاستقلال فإننا نستخدم الاختبار نفسه لكليهما ويتم هذا الاختبار أو الا بحساب التكرارات المتوقعة E_{ij} للخلية (i, j) حيث

$$i = 1, 2 ,..., r \quad , \quad j = 1, 2 ,..., c$$

وتعطى E_{ij} تحت صحة فرض العـدم H_0 أن المعياريـن A, B مستقلان بالعلاقـة التالية:

$$E_{ij} = \frac{O_{i.}(O_{.j})}{O_{..}} , \;\; i = 1,2,...., r \;\; ; \; j = 1,2,..., c$$

ومن السهل برهان العلاقات التالية:

$$\sum_{i}^{r} \sum_{i}^{c} E_{ij} = n$$

$$\sum_{i}^{r} E_{ij} = \sum_{i}^{r} O_{ij} = O_{.j} \quad , \; j = 1,2,\ldots, c$$

$$\sum_{j}^{c} E_{ij} = \sum_{j}^{c} O_{ij} = O_{i.} \quad , \; i = 1,2,\ldots, r$$

ثم نحسب الإحصاءة χ_o^2 من العلاقة التالية:

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \qquad (٨,٤)$$

تحت صحة فرض العدم H_0 فإن الإحصاءة χ_0^2 تتوزع تقريبا كتوزيع مربع كاي بدرجات حرية $\upsilon = (r-1)(c-1)$ عند مستوى معنوية α يمكن تحديد فترة قبول H_0 بالفترة $\left(0, \chi^2_{\upsilon,\alpha}\right)$ ونرفض H_0 عندما تكون χ_o^2 المحسوبة خارج هذه الفترة.

مثال (٨,٧)

اخذت عينة تتكون من 360 طالب في كلية العلوم وجرى تقسيمهم حسب مقدرتهم في مادتي الرياضيات والإحصاء حسب جدول (٨,١٢) كما يلي:

جدول (٨,١٢)

المقدرة في الإحصاء	المقدرة في الرياضيات			المجموع
	منخفضة	متوسطة	عالية	
منخفضة	63	42	15	120
متوسطة	58	61	31	150
عالية	14	47	29	90
المجموع	135	150	75	360

والمطلوب اختبار عند مستوى معنويـة $\alpha = 0.01$ فيما إذا كانت مقدرة الطالب في الإحصاء مستقلة عن مقدرته في الرياضيات.

الحل

أولا: صياغة الفرض الإحصائي.

H_0 : مقدرة الطالب في الإحصاء مستقلة عن مقدرته في الرياضيات

H_1 : مقدرة الطالب في الإحصاء تعتمد على مقدرته في الرياضيات

ثانيا: اختيار الإحصاءة ثم حسابها.

نختار الإحصاءة χ_0^2 حيث

$$\chi_0^2 = \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad , \; i = 1,\ 2,\ 3 \;\; , \;\; j = 1,\ 2,\ 3$$

ثم نحسب E_{ij} لكل خلية كما يلي

$$E_{ij} = \frac{O_{i.}\,(O_{.j})}{O..} \quad , \; i = 1,2,3, \; ; \; j = 1,2,3$$

كما يلي:

$$E_{12} = \frac{41 \times 82}{376} = 8.94 \;\; ; \;\; E_{22} = \frac{65 \times 82}{376} = 14.18 \;\; ; \;\; E_{23} = \frac{270 \times 82}{376} = 58.88$$

$$E_{21} = \frac{135 \times 150}{360} = 56.25 \;\; ; \;\; E_{22} = \frac{150 \times 150}{360} = 62.5 \;\; ; \;\; E_{23} = \frac{75 \times 150}{360} = 31.25$$

$$E_{31} = \frac{135 \times 90}{360} = 33.75 \;\; ; \;\; E_{32} = \frac{150 \times 90}{360} = 37.5 \;\; ; \;\; E_{33} = \frac{75 \times 90}{360} = 18.75$$

ويمكن وضع التكرارات المشاهدة O_{ij} في أعلى الركن الأيسر للخلية (i,j) ونوضح ذلك في جدول (٨،١٣) كما يلي:

جدول (١٣ , ٨)

المقدرة في الاحصاء	المقدرة في الرياضيات		
	منخفضة	متوسطة	عالية
منخفضة	62 (45.0)	42 (50.0)	15 (25.0)
متوسطة	58 (56.25)	61 (62.5)	31 (31.25)
عاليـــة	14 (33.75)	47 (37.5)	29 (13.75)

ودرجات الحرية υ هي $\upsilon = (3-1)(3-1) = 4$

ثم نحسب χ_0^2 من جدول (١٣ , ٨) كما يلي:

$$\chi_0^2 = \frac{(62-42.0)^2}{42.0} + \frac{(58-56.25)^2}{56.25} + \cdots + \frac{(29-13.75)^2}{13.75} = 32.14$$

ثالثا: تحديد منطقة قبول ورفض H_0.

تحدد منطقة قبول H_0 بالفتـرة $\left(0, \chi^2_{\upsilon,\alpha}\right)$ حيث في مثالنا هي $\left(0,\ \chi^2_{4,\ 0.01}\right)$ وحيث قيمة $\chi^2_{4,\ 0.01} = 13.277$ من جدول مربع كاي (٣) آخر الكتاب وبالتالي فإن فترة قبول H_0 هي $(0, 13.277)$ ومنطقة رفض H_0 خارج هذه الفترة.

رابعا: القرار الإحصائي.

حيث $\chi_0^2 = 32.14$ المحسوبة واقعة خارج فترة قبول H_0 والتي هي $(0, 13.277)$ فإننا نرفض H_0 القائل أن المقدرة في الإحصاء مستقلة عن المقدرة في الرياضيات ونقبل الفرض البديل H_1 أي غير مستقلة.

مثال (٨,٨)

الجدول (٨,١٤) يمثل نتيجة فحص أجهزة من ثلاث مؤسسات مختلفة إلى جهة ما. المطلوب اختبار عند مستوى معنوية $\alpha = 0.01$ فيما إذا كانت نوعية الاجهزة من المؤسسات الثلاث لها نفس الجودة (أي متجانسة).

جدول (٨,١٤)

المؤسسات	نوعية الأجهزة			المجموع
	مرفوضة	مقبولة	ممتازة	
المؤسسة الأولى	12	23	89	124
المؤسسة الثانية	8	12	62	82
المؤسسة الثالثة	21	30	119	170
المجموع	41	65	270	376

الحل

أولا: صياغة الفرض الإحصائي.

نوعية الأجهزة للمؤسسات الثلاثة متجانسة: H_0

نوعية الأجهزة للمؤسسات الثلاث غير متجانسة: H_1

ثانيا: اختيار الإحصاءة ثم حسابها.

نختار الإحصاءة χ_0^2 حيث

$$\chi_0^2 = \sum_{i=1}^{3} \sum_{i=1}^{3} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \quad i = 1, 2, 3, \quad j = 1, 2, 3$$

ودرجات الحرية υ هي $\upsilon = (3-1)(3-1) = 4$ وتحسب قيم E_{ij} من جدول (٨,١٤) كما يلي:

$$E_{11} = \frac{41\times124}{376} = 13.52 \;;\; E_{12} = \frac{65\times124}{376} = 21.44 \;;\; E_{13} = \frac{270\times124}{376} = 89.04$$

$$E_{12} = \frac{41\times82}{376} = 8.94 \;;\; E_{22} = \frac{65\times82}{376} = 14.18 \;;\; E_{23} = \frac{270\times82}{376} = 58.88$$

$$E_{13} = \frac{41\times170}{376} = 18.54 \;;\; E_{23} = \frac{65\times170}{376} = 29.39 \;;\; E_{33} = \frac{270\times170}{376} = 122.07$$

ويمكن تلخيص القيم O_{ij} في جدول (٨,١٤) والقيم المتوقعة E_{ij} في أسفل الركن الأيمن للخلية (i, j) كما يلي:

جدول (٨,١٥)

المؤسسات	نوعية الأجهزة		
	ممتازة	مقبولة	مرفوضة
الأولى	89 (89.04)	23 (21.44)	12 (13.52)
الثانية	62 (58.88)	12 (14.18)	8 (8.94)
الثالثة	119 (122.07)	30 (29.39)	21 (18.54)

ثم نحسب χ_0^2 من جدول (٨,١٥) السابق كما يلي:

$$\chi_0^2 = \frac{(12-13.52)^2}{13.52} + \frac{(8-8.94)^2}{8.94} +\ldots+ \frac{(119-122.07)^2}{122.07} = 1.3$$

ثالثا: تحديد منطقة قبول ورفض H_0.

تحدد منطقة قبول H_0 بالفترة $(0, \chi_{v,\alpha})$ وفي مثالنا هي الفترة $(0, \chi^2_{4,\ 0.01})$ وأن

قيمة $\chi^2_{4,0.01} = 13.28$ تستخرج من جدول مربع كاي (٣) آخر الكتاب وبذلك فإن فترة قبول H_0 هي (0, 13.28) ومنطقة رفض H_0 خارج هذه الفترة.

رابعا: القرار الإحصائي

حيث إن قيمة $\chi^2_0 = 1.3$ المحسوبة واقعة داخل فترة قبول H_0 (0, 13.28) فإننا لا نستطيع رفض H_0 أي أن نوعية الأجهزة متجانسة للمؤسسات الثلاث.

ملاحظة

إذا أخذت جداول التوافق الرتبة (2 x 2) أي أن كل من المعيارين A, B لا تنقسم إلا إلى قسمين فقط أو صفتين فقط في دراسة العلاقة المعنوية، فإننا نطلق على جدول التوافق في هذه الحالة بجدول الاقتران (contingency table)، ويمكن اعتباره حالة خاصة من جدول التوافق وله درجات حرية تساوي الوحدة حيث إن معرفة تكرار واحد فيه يمكن معرفة باقي التكرارات (الثلاث الباقية) في ظل معرفة المجموع الكلي ومجموع كل من خلايا الصف أو خلايا العمود. ويمكن حساب الإحصاءة χ^2_0 من العلاقة (٨, ٤) السابقة أو بالعلاقة (٨, ٥) التالية والمستنتجة من جدول الاقتران (٨, ١٦) التالي:

جدول (٨, ١٦)

المعيار A	المعيار B	
	الصفة الثانية	الصفة الأولى
خاصية I	a	b
خاصية II	c	d

$$\chi^2_1 = \frac{n(ad - cb)^2}{2(c+b)(c+d)(a+c)(b+d)} \qquad (٨, ٥)$$

ونوضح ذلك بالمثال التالي.

مثال (٩ ,٨)

في عينة عشوائية من مائة عامل من عمال المصانع وجد من بين 75 عامل متعلم 20 عامل يدخن ومن العمال غير المتعلمين 6 عمال لا يدخنون. فهل يمكن القول إن معيار التدخين مستقل عن معيار التعليم. عند مستوى معنوية $\alpha = 0.05$.

الحل

أولا: صياغة الفرض الإحصائي.

التدخين مستقل عن التعليم: H_0

التدخين غير مستقل عن التعليم: H_1

$\alpha=0.05$

ثانيا: اختيار الإحصاءة ثم حسابها.

نحسب الإحصاءة χ_0^2 بطريقتين:

(ا) باستخدام العلاقة (٤ ,٨) وهي

$$\chi_0^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

ويلزم ذلك حساب التكرارات المتوقعة E_{ij} حيث يشمل الجدول (١٧ ,٨) على التكرارات المشاهدة كما يلي:

جدول (١٧ ,٨)

التعليم / التدخين	متعلم	غير متعلم	المجموع
يدخن	55	19	74
لا يدخن	20	6	26
المجموع	75	25	100

التكرارات المتوقعة هي

$$E_{11} = \frac{75(74)}{100} = 55.5 \quad ; \quad E_{12} = \frac{25(74)}{100} = 18.5$$

$$E_{21} = \frac{75(26)}{100} = 19.5 \quad ; \quad E_{22} = \frac{25(26)}{100} = 6.5$$

ويمكن تكوين جدول (٨,١٨) يشتمل على التكرار المشاهد O_{ij} في أعلى الركن الأيسر للخلية (i, j) والتكرار المتوقع E_{ij} في أسفل الركن الأيمن لنفس الخلية (i, j) كما يلي:

جدول (٨,١٨)

التعليم / التدخين	متعلم	غير متعلم
يدخن	55 (55.5)	19 (18.5)
لا يدخن	20 (19.5)	6 (6.5)

ونحسب χ_0^2 من جدول (٨,١٨) كما يلي:

$$\chi_0^2 = \frac{(55-55.5)^2}{55.5} + \frac{(20-19.5)^2}{19.5} + \frac{(19-18.5)^2}{18.5} + \frac{(6-6.5)^2}{6.5}$$

$$= 0.005 + 0.013 + 0.014 + 0.038 = 0.07$$

(ب) باستخدام العلاقة (٨,٥) والجدول (٨,١٦) كما يلي:

$$\chi_1^2 = \frac{n(ab-cd)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{100(55\times 6 - 19\times 20)^2}{75(25)(74)(26)} = 0.07$$

وهي نفس قيمة χ_0^2 المحسوبة في (ا).

ثالثا: تحديد منطقة قبول ورفض H_0.

تحدد منطقة قبول H_0 بالفترة $\left(0,\ \chi^2_{\upsilon,\alpha}\right)$ حيث في مثالنا هي الفترة $(0, \chi^2_{1, 0.05})$ وباستخدام جدول مربع كاي (٣) آخر الكتاب فإن فترة قبول H_0 هي $(0, 3.841)$.

رابعا: القرار الإحصائي.

وحيث إن قيمة $\chi_0^2 = 0.07$ المحسوبة واقعة داخل فترة قبول H_0 $(0, 3.841)$ فإننا لا نستطيع رفض H_0 أي أن ظاهرة التدخين مستقلة عن التعليم.

(٨, ٤) تمارين

١ - في تجربة لرمي حجر النرد ستين مرة سجلت نتائجها في الجدول التالي:

ظهور الوجه	1	2	3	4	5	6
التكرار المشاهد	15	7	4	11	6	17

فإذا كان رامي حجر النرد أمينا وغير متحيز. المطلوب اختبار أن حجر النرد متزن عند مستوى معنوية $\alpha = 0.01$.

٢ - في تجربة على سلالات زهور من نوع معين، حصل أحد الباحثين على 120 زهرة قرمزية ذات مبسم أخضر، 40 زهرة قرمزية ذات مبسم أحمر، 36 زهرة حمراء ذات مبسم أخضر و 13 زهرة حمراء ذات مبسم أحمر. وطبقا لنظرية مندل للوراثة أن هذه الأصناف لا بد من الحصول عليها طبقا للنسب $9:3:3:1$.

فهل نتائج التجربة تتفق مع النظريـة عند مستـوى معنوية $\alpha = 0.05$.

٣ - الجدول التالي يمثل تصنيف حوادث السيارات في مجتمع ما على حسـب عمرالسائق وذلك للسائقين الذين تقل أعمارهم عن 25 عاما .

فئات السن	15–16	17–18	19–20	21–22	23–24
التكرار المشاهد (عدد السائقين)	5	12	10	8	5

والنسب المتوقعة في هذه المجموعات لهذا المجتمع هي تقريبا %10 ، %20، %20 ، %25 و %25 على التوالي والمطلوب اختبار ما إذا كانت مـعـدلات الحـوادث للسـائقين تحت 25 عـاما لها النسب نفســها عند مستوى معنوية $\alpha = 0.05$.

٤ - عينة مـن 120 طالبا تم تقسيمها حسب أوزانهم ودرجاتهم في الامتـحـان حسب الجدول التالي والمطلوب اختبار الفرض القائل أنه لا توجد علاقة بين الوزن والتحصيل العلمي عند مستوى معنوية $\alpha = 0.05$ و $\alpha = 0.01$.

التقدير / الوزن	ممتاز	جيـد	متوسط	ضعيف
نحيف	14	11	10	5
متوسط الوزن	10	16	16	14
بدين	3	4	7	10

٥ - عدد الأفراد المنتمين لجنس ما لا بد أن يقسموا بالنسب 0.16 ، 0.48 ، 0.16 و 0.20 طبقا لفصائل الدم الأربع . فإذا كان لدينا التكرارات المشاهدة لجنس آخر كما يلي : 100 ، 130 ، 360 و 180 . المطلوب اختبار ما إذا كان لها نفـس توزيع أنواع الدم .

٦ - البيانات التالية تمثل نتائج بحث لتحديد توزيع النوع لأطفال 32 عائلة لدى كل منها أربعة أطفال. استخدم توزيع ذي الحدين لحساب التكرارات المتوقعة بفرض أن $p=0.5$ و $n=4$. بعد ذلك طبق اختبار χ^2 لنرى ما إذا كان نموذج توزيع ذي الحدين مناسب لهذه المشكلة.

عدد الأولاد الذكور	0	1	2	3	4
عدد الأسر	4	10	8	7	3

٧ - لمعرفة تأثير دواء معين يعتقد أنه مؤثر على علاج البرد أجريت تجربة على 164 مصابا بالبرد تناول نصفهم هذا العقار وتناول النصف الآخر أقراص السكر وكانت نتيجة العلاج كالموضحة في الجدول التالي. اختبر الفرض بأن العقار وأقراص السكر لهما التأثير نفسه. عند مستوى معنوية $\alpha = 0.05$.

	تم علاجهم	اكثر معاناة	لا تأثير
العقار	50	10	22
أقراص السكر	44	12	26

٨ - إذا علم أن متوسط عدد المكالمات اليومية التي تقوم بها ربة البيت في بلد ما ثلاث مكالمات يوميا فمن عينة عشوائية من ألف سيدة سجلت المشاهدات التالية

عدد المكالمات	0	1	2	3	4	5	6
عدد السيدات	150	232	551	43	17	4	3

فهل تتفق هذه المعلومـات مع الرأي القائل بأنها مأخوذة من توزيع بواسـون عند مستوى معنوية $\alpha = 0.05$.

٩ - عينة عشوائية من 126 عامل من عمال أحد المصانع وجد أن أجورهم اليومية تتوزع على النحو التالي:

فئات الأجر اليومي	100 –	110 –	120 –	130 –	140 –	150 –	160–170
عدد العمال	8	40	20	27	15	9	7

المطـلوب اختبار جودة التوفيق بأن هذه البيانات المشاهدة تتبع التوزيع الطبيعي عند مستوى معنوية $\alpha = 0.05$.

١٠ - لدراسة العلاقة بين وقت العمل وإنتاجية العامل أخذت عينة عشوائية مـن ألفي عامل وسجلت المشاهدات كما في الجدول التالي:

وقت العمل / الإنتاجية	صباحا	ظهرا	مساء	المجموع
100–110	125	447	342	914
110–120	151	98	116	365
120–130	223	361	187	721
المجموع	499	856	645	2000

اختبر الفرض القائل بأن وقت العمل مستقل عن الإنتاجية. (**ملحوظة:** الإنتاجية بالوحدة خلال فترة العمل وهي ثماني ساعات لكل فترة).

١١ - لدراسة العلاقة بين لون البشرة ولون الشعر، تم اختيـار 100 شخص وتم تصنيفهم في الجدول التالي:

لون الشعر / لون البشرة	أصفر	أسود
أبيـض	32	20
أسـود	13	35

هل توجد علاقة بين لون الشعر ولون البشرة ؟ استخدم مستوى معنــوية $\alpha = 0.01$.

١٢ - يرغب بعض الباحثين دراسة العلاقة بين فصيلة الدم والإصابة بقرحة الأمعاء. تم اختيار عينة حجمهــا 1301 مريضا بقرحة الأمعاء وعينة أخرى حجمهــا 6313 شخصا سليما من قرحة الامعاء وتم تصنيف مفردات العينتين حسب فصيلة دم كل شخص فحصلنا على الجدول التالي :

المجموع	نوع الشخص		فصيلة الدم
	سليـم	مريض	
3590	2892	698	O
3097	2625	472	A
672	570	102	B
255	226	29	AB
7614	6313	1301	المجموع

هل يمكن القول إن هناك علاقة بين نوع فصيلة الدم والإصابة بقرحة الأمعاء ؟ استخدم مستوى معنوية $\alpha = 0.05$.

١٣ - ادّعى شخص أن النسبة بين أعداد ثلاثة أنواع من السيارات الموجــودة فــي المملكة هــي 6 : 3 : 1، اخترنا 100 سيارة من هذه الأنواع الثلاثة عشوائــيا

فوجدنا أعداد هذه الأنواع في هذه العينة هي 10,40,50 على الترتيب استخدم إختبار مربع كاي لاختبار إدعاء هذا الشخص عند مستوى معنوية $\alpha = 0.01$.

١٤ - اختير 100 طالب بطريقة عشوائية وصنفوا حسب التخصص والتقدير في إحدى المواد فكان توزيع الطلاب كما هو بالجدول التالي:

التقدير / التخصص	A	B	C	D	F
الأول	8	7	20	10	5
الثاني	1	2	5	10	12
الثالث	1	1	5	10	3
المجموع	10	10	30	30	20

(i) أختبر استقلالية تقدير الطلاب عن تخصصاتهم عند مستوى معنوية $\alpha = 0.05$.

(ii) إذا اخترنا 50 طالبا من التخصص الأول و 30 طالبا من التخصص الثاني و 20طالباِ من التخصص الثالث وصنفنا الطلاب حسب تقديراتهم فحصلنا على نفس الجدول السابق فما هو اسم الاختبار في هذه الحالة وكيف يصاغ فرض العدم H_0 في هذه الحالة وكذلك الفرض البديل H_1 وضح ذلك.

الفصل التاسع

أساسيات تصميم التجارب
Principles of Experimental

● مقدمة ● التجريب والتجربة ● القواعد الأساسية لتصميم التجـارب ● أهداف التجربـة ● الوحدة التجريبية والمعالجة ● الخطأ التجريبي ● المكررات ووظائفها ● التحكم في الخطأ ● التعشية (العشوائية) ● الخطوات الواجب اتباعها لاستخدام الأسلـوب الإحصائـي ● محاذير عامـة ● أنواع تصميمـات التجارب ● تمارين

(١ , ٩) مقدمة

Introduction

نشأ علم تصميم التجارب في محطات التجارب الزراعية وتطور فـيـهـا. يخطىء من يعتقد أن إقامة التجارب دون الإلمام بالأسس الإحصائية أمر سليم، كما يخطىء من يعتقد أن الإلمام بالإحصاء دون الإحاطة بمادة الدراسة أمر سليم أيضا، وسلامة تصميم أي تجربة أمر ضروري لسلامة النتائج واستواء التفسير. إذ يصبح التفسير خطأ إذا تم على أساس خاطيء. وتجري التجربة لاختبار فرض معين نحو أحد العوامل أو بعض العوامل المتغيرة مع تثبيت العوامل الأخرى خلاف عامل أو عوامل الدراسة. ويلعب الإحصاء دورا مهما في تقدم التجارب أو يساعد عـلـى وضع التصميمـات الملائمة لكل نوع من أنواع التجارب، وفي تحليل النتـائـج و تفسيرها تفسيرا إحصائيا. ويغطي هذا الفصل بعض الموضوعات العامة والرئيسة

في تصميم التجارب وتشمل تعريف التجربة وبعض أنواعها والقواعد الأساسية لتصميم التجربة، وأهداف التجربة، وتعريف الوحدة التجريبية والمعالجة. كذلك يتم عرض بعض الموضوعات المتعلقة بأساسيات تصميم التجارب كالتعشية والتكرار وخطوات التجربة.

(٩,٢) التجريب والتجربة

Experimentation

تقدمت الطرق التجريبية في كثير من ميادين العلوم، والتجربة (experiment) هي الوسيلة المباشرة والفعالة في الحصول على حقائق حيثما يمكن التجريب. وهناك عدة تعاريف معروفة للتجربة منها أنها تساؤل مخطط للحصول على حقائق جديدة أو لتأكيد أو لمعارضة نتائج تجارب سابقة. وتختصر التجربة الوقت والجهد للحصول على تلك الحقائق أو الإجابات وتتعدد أنواع التجارب فمنها:

(٩,٢,١) التجربة التمهيدية

يرغب الباحث في التجربة التمهيدية (preliminary experiment) في الاطمئنان إلى العوامل التي نقوم بدراستها، ومن ذلك النوع من التجارب ما يلي:

(٩,٢١,١) التجربة الاستطلاعية

تجري التجربة الاستطلاعية (pilot experiment) ليطمئن الباحث على العوامل تحت الدراسة قبل إجراء التجربة الضابطة (controlled experiment) وعلى سبيل المثال لو حصلنا على بذور مستوردة لنبات ما لا نعرف مواعيد زراعتها بالتحديد. فنجري تجربة استطلاعية على مواعيد متسعة حتى يمكن تحديد ميعاد تقريبي للزراعة.

(٩,٢,١,٢) التجربة الانتقائية

يجري الباحث التجربة الانتقائية (screening experiment) للاستفادة من بعض

مواد الدراسة. فمثلا يرغب في اختبار مجموعة من المبيدات الحشرية لمعرفة تأثيرها على الحشرات. يجري الباحث تجربة لانتقاءبعض من هذه المبيدات الحشرية.

(٩,٢,١,٣) التجربة التخطيطية

عند اختيار مستويات معينة لعامل ما، يجري الباحث التجربة التخطيطية (sighting experiment) لاختبار مستويات العامل، فمثلا اختبار مستويات تسميد، درجات حرارة، مواعيد زراعة أو مستويات ضغط جوي.

(٩,٢,٢) التجربة الحاسمة

يقصد بالتجربة الحاسمة (crucial experiment) تلك التجربة التي تمكننا من معرفة الأسباب المحدثة لظاهرة ما. فمثلا في الدراسات الأحيائية (البيولوجية) في أمراض النبات نجري التجربة الحاسمة لمعرفة الفطريات المسببة لمرض ما وذلك مثل إجراء التجربة الضابطة.

(٩,٢,٣) التجربة الضابطة

التجربة الضابطة (controlled experimnet) هي تلك التجربة التي تجرى للحصول على البيانات وتجرى بعد إجراء التجارب التمهيدية وتثبيت جميع العوامل عدا العامل تحت الدراسة. إذا تضمنت التجربة عاملا واحدا تسمى **تجربة بسيطة**، أما إذا تضمنت عدة عوامل فتسمى **تجربة عاملية**. يمكن النظر إلى التجارب على أنها تشمل ثلاثة أقسام التمهيدية فالحاسمة (التخطيطية) ثم الضابطة وكلها تؤدي إلى الأخرى. يحاول الباحث عدد كبير من المعالجات في التجربة التمهيدية للحصول على بعض منها للتجارب المستقبلية، ويلاحظ أنه من الضروري في حالة التجارب الحاسمة والضابطة تعريف المجتمع الذي ستعمم عليه النتائج وتصمم التجربة بناء على ذلك. في النهاية فإن كل تجربة ماهي إلا خطة محكمة لتمد الباحث بإجابات لواحد أو أكثر من الأسئلة، وتحدد للباحث المقارنات التي يريدها والتي تساعد في

تصميم التجربة. ويلعب الإحصاء دورا مهما في تقدم التجارب إذ يساعد على وضع التصميمات الملائمة لكل نوع من أنواع التجارب وتفسيرها تفسيرا إحصائيا. يساعد الإحصاء أيضا في تعميم النتائج باحتمالات معينة.

فإن التجربة هي مجموعة القواعد التي يستخدمها الإحصائي لسحب عينة من المجتمع، وهنا تبرز أهمية تعريف المجتمع، وتلك المجموعة من القواعد هي تصميم التجارب. على سبيل المثال، استخدام المشاهدات المزدوجة وغير المزدوجة تعتبر تصميم التجربة لتجارب مقارنة معالجتين.

(٩,٣) القواعد الأساسية لتصميم التجارب
Principal Steps of Experimental Design

يعتبر السير رونالد فيشر (Fisher) مبتكرا استخدام الإحصاء في تصميم التجارب ولعدة سنوات كان مسئولا عن الإحصاء وتحليل البيانات بمحطة روثهمتد للتجارب الزراعية بإنجلترا (1933). وقدم كثيرون إضافات معنوية بعد ذلك منهم يتس (Yates)، بوكس (Box)، كمبثورن (Kampthorne)، وكوكران (Cochran). وكان الاستخدام الرئيسي لتصميم التجارب في الأبحاث الزراعية والأحيائية (البيولوجية) أما الآن فإن تصميم التجارب يستخدم في كثير من المجالات العلمية المختلفة بخلاف الزراعة والعلوم الأحيائية (البيولوجية) وعلوم الطب وعلوم الهندسة والعلوم السلوكية والعلوم الاجتماعية و كثير من المجالات التي يمكن قبول تصميم التجارب الإحصائي فيها وتحليل البيانات.

يستخدم التجريب بواسطة جميع الباحثين في جميع المجالات وقبل أن نتكلم عن أساسيات تصميم التجارب سنورد المثال التالي.

مثال (٩,١)

أفترض أن أحد مهندسي المعادن أراد أن يدرس الفرق بين تأثير طريقتين مختلفتين لتقسية المعادن (الأولى الغمس في الزيت والثانية الغمس في محلول ماء

ملحي) على سبيكة معينة من الألومنيوم وكان هدف التجربة معرفة أي الطرق التى يؤدي إلى أعلى تقسية للسبيكة.

قام المهندس بتخصيص عدد من القضبان المصنوعة من السبيكة لكل نوع من أوساط المحاليل ثم يقيس التقسية بعد غمسها في المحلول ثم حساب الوسط الحسابي للتقسية للقضبان لكل محلول لتحديد أي وسيلة تعطي تقسية أكبر من الأخرى.

يتبادر إلى الذهن من تلك التجربة مجموعة من الأسئلة نوردها فيما يلي:

- هل نهتم بهذين المحلولين فقط من وسائط الغمس المختلفة؟
- هل نريد دراسة عوامل أخرى تؤثر على التقسية؟ وهل هناك عوامل نريد تقديرها؟
- ما عدد القضبان التي نغمسها في كل محلول؟
- كيف يمكن تخصيص القضبان للمحاليل المختلفة؟ وبأي ترتيب يمكن جمع البيانات؟
- ما هي طرق تحليل البيانات التي يجب استخدامها؟
- ما هو الفرق بين المتوسطات الذي يمكن اعتباره مهما؟

وقد يكون هناك مزيدا من الأسئلة يراد الإجابة عليها قبل التجربة. ويجب أن نعلم أن النتائج والقرارات التي تعمم تعتمد بدرجة كبيرة على طريقة جمع البيانات ولتوضيح ذلك، افترض أن المهندس استخدم مجموعة قضبان المحلول الأول من فرن ما وقضبان المحلول الثاني من فرن أخرى. لا يستطيع المهندس في هذه الحالة معرفة مصدر الفرق بين المتوسطين هل هو راجع للفرن أم راجع للمحلول. و يدفعنا ذلك إلى الاستعانة بإحصائي عند تصميم التجربة حتى يكون على بينة من أن البيانات التي ستجمع هي التي نريدها وسيتم تحليلها بطريقة مناسبة. وهناك محورين لأي تجربة وهما التصميم والتحليل وهما مرتبطان ببعضهما البعض فالتحليل يعتمد على التصميم المستخدم في التجربة. وهناك عوامل أساسية لكل تجربة وهذه العوامل هي التكرار والتحكم في الخطأ والعشوائية. وسوف نقوم

بشرحها مع بعض المصطلحات الأخرى المهمة في التجربة في الأقسام التالية.

(٩,٤) أهداف التجربة

The Aims of the Experiment

عند تصميم أي تجربة فإن تحديد الأهداف بوضوح هو تلك الأسئلة المراد الإجابة عليها والفروض المراد اختبارها والتأثيرات المراد تقديرها. ومن المهم أن يتم تقسيم الأهداف حسب أهميتها إلى أهداف أساسية وأخرى فرعية حيث إن بعض التصميمات تعطي دقة أكبر لبعض المقارنات بين المتوسطات عن البعض الآخر. إذا عرفنا كمية المعلومات (I) بأنها مقلوب تباين المتوسط$(\frac{\sigma^2}{n})$ فإن$I = \frac{n}{\sigma^2}$ وواضح أن I تقل كلما زادتσ^2 وتزيد كلما زادت n. وتؤدي زيادة n أيضا إلى زيادة حساسية المقارنة بين متوسطي عينتين بمعنى كشف أقل فرق ممكن بين متوسطي عينتين بمعنى كشف أقل فرق ممكن بين متوسطي مجتمعين كلما زاد حجم العينة. وهذا بدوره يوضح أنه من المهم جدا أن نعرف المجتمع الإحصائي الذي ستعمم عليه نتائج الاستدلال الإحصائي والذي أيضا ستسحب منه العينة عشوائيا. ففي مثال (٩,١) السابق إذا تعددت الأفران التي يأخذ منها الباحث القضبان فإنه من الخطأ عمل أي استدلال إحصائي من تجربة بهذا الشكل، ولتحاشي هذا يجب أن يشمل التصميم العاملين معا وسيط المحلول وأيضا الأفران المستخدمة.

(٩,٥) الوحدة التجريبية والمعالجة

Experimental Unit

تعرف الوحدة التجريبية على أنها جزء من المادة التجريبية والتي تطبق عليها معالجة واحدة. أما المعالجة (treatment) فتعرف على أنها طريقة التأثير في الوحدة التجريبية وهي الأسلوب الذي يراد أن يقاس تأثيره ومقارنته بالمعالجات

الأخرى. وقد تكون الوحدة التجريبية حيوان أو عشرة طيور في قفص أو نصف ورقة نبات أو قضيب من سبيكة الألمونيوم كما في مثال (٩،١) السابق. أما المعالجة فقد تكون طريقة زراعية، نوع سماد، مستويات رطوبة، مستويات حرارة مختلفة، طرق تقسية معادن كما في مثال (٩،١) السابق... الخ. ويقاس تأثير المعالجة على ما يسمى بوحدة المعاينة (sampling unit) وتعرف على أنها نسبة من الوحدة التجريبية وقد يكون كل الوحدة التجريبية حيوانا أو نسبة منها كطائر من ضمن عشرة طيور في قفص. ويجب عند اختيار مجموعة من المعالجات تعريف كل معالجة بعناية وخاصة بالنسبة للمعالجات الأخرى حتى يضمن إجابة الأسئلة المتعلقة بأهداف التجربة.

(٩،٦) الخطأ التجريبي
Experimental Error

إن أحد صفات المواد التجريبية الاختلاف ويعرف الخطأ التجريبي على أنه مقياس للاختلاف الموجود بين مشاهدات الوحدات التجريبية والمعالجة بنفس المعالجة. فمثلا، إذا كان لدينا عشرة طيور في قفص تم تغذيتها بغذاء معين فإن الوحدة التجريبية هنا هي القفص، ولذلك فإننا نريد أقفاص أخرى كل منها به عشرة طيور قبل أن نقيس الاختلاف بين الوحدات المعالجة بنفس المعالجة.

يأتي الاختلاف من مصدرين رئيسين: **الأول** الاختلاف الموجود في طبيعة المواد التجريبية منها والتي ستطبق عليها المعالجات **والثاني** الناتج من الوقوع في أخطاء مثل فقدان الانتظام أو التماثل في بعض الظروف عند تطبيق التجربة. وكمثال على المصدر الأول وجود اختلافات وراثية بين مجموعة من طيور قفص التجارب. أو إحدى تجارب التغذية ولم تؤخذ ذلك في الاعتبار الأول. وإذا وضعت الأقفاص التي بها الطيور في أماكن مختلفة الحرارة والضوء أو أي عوامل أخرى ليست داخلة في التجربة فهذا مثال على فقدان الانتظام لطرق التجربة. ويختلف الحجم النسبي لكلا من المصدرين باختلاف التجربة وطريقة البحث وحقل الدراسة.

تعتمد فترة الثقة وقوة الاختبار الإحصائي في النهاية على تباين المتوسط $V(\overline{X}) = \frac{\sigma^2}{n}$ ولذا للحصول على فترات ثقة ضعيفة أو قوة عالية فهناك نقطتين رئيسيتين للدراسة. بناء على ذلك فإنه يجب زيادة الجهود الممكنة لتقليص حجم الخطأ التجريبي بغرض تحسين قوة الاختبار وتقليل حجم فترات الثقة أو تحقيق أية أهداف أخرى مرغوبة، وهذا يمكن أن يتحقق بمعالجة مصدري الخطأ التجريبي كما يلي:

- معاملة المواد التجريبية بطريقة تقلل من الاختلافات الموجودة في طبيعتها (المصدر الأول).
- إدخال تحسينات على الأسلوب التجريبي وسوف نناقش ذلك في البنود التالية.

(٩,٧) المكررات ووظائفها

Replication Role

يقصد بالمكررات (replications) إعادة المعالجة حتى تظهر مرتين أو أكثر في التجربة أي أنه حينما تظهر المعالجة أكثر من مرة في التجربة نقول إنها مكررة (replicated) وتعتبر المكررات من أهم التحسينات التي حدثت في تصميم التجارب في الحقبة الماضية.

(٩,٧,١) وظائف المكررات

وأهمها هي:

- مد المجرب بتقدير الخطأ التجريبي.
- تحسين دقة التجربة بإقلال الانحراف المعياري لمتوسط المعالجة (الخطأ التجريبي).
- زيادة مجال الاستدلال الإحصائي للتجربة وذلك بالاستخدام المناسب لوحدات تجريبية أكثر اختلافا.

- التأثير في ضبط تباين الخطأ.

وسنتناول هذه الوظائف للمكررات بالتفصيل كما يلي:

(٩,٧,١,١) تقدير الخطأ التجريبي

إن تقدير الخطأ التجريبي مطلوب لاختبارات المعنوية وتقديرات فترات الثقة. وتسمى التجربة التي تظهر فيها المعالجة مرة واحدة تجربة من مكرر واحد وليس هناك تقدير للخطأ التجريبي من تلك التجربة فمن المستحيل تقدير التباين من وحدة واحدة، ولا يعتبر أيضا أحد أكثر من مرة واحدة على نفس الوحدة التجريبية مكررا. وهنا ليس من الممكن شرح الاختلاف المشاهد بين الوحدات التجريبية وليس هناك أي طريقة موضوعية لمعرفة هل هو اختلاف بين المعالجات أو بين الوحدات التجريبية؟ أي أنه في حالة عدم وجود تقدير للخطأ التجريبي فليس هناك أي طريقة لمعرفة أن الاختلاف المشاهد اختلاف حقيقي أو راجع للاختلاف الموجود في طبيعة الوحدات التجريبية.

ملحوظة

بعض التجارب الخاصة العاملية من مكرر واحد أو جزء من مكرر شاملة عدد كبير من العوامل أو أنواع المعالجات ممكن أن تستخدم وتمدنا بتقدير للخطأ التجريبي وذلك عند توافر فروض معينة.

(٩,٧,١,٢) تحسين دقة التجربة

إن تقدير متوسط المجتمع مثل متوسط المعالجة فالمشاهد يصبح أكثر دقة حينما يزداد عدد المكررات. وإذا تمكنا من كشف فرق بين متوسطين من خمس وحدات بمعنى رؤية الاختلاف بينهما معنويا مستخدما أربعة مكررات فإنه في تجربة من 16 مكررا لنفس المتوسطين سيكشف فرقا قدره 0.5 وذلك لأن نسبة الخطأ

المعياري ستكون $\left[\frac{\sigma}{\sqrt{4}} : \frac{\sigma}{\sqrt{16}}\right]$ أي بنسبة $1:2$ لاحظ أن زيادة عدد المكررات قد يدفعنا لاستخدام وحدات تجريبية أقل تجانسا أو أسلوب أقل عناية وهذا يعطي مصدر مجتمع جديد بخطأ تجريبي كبير ومع ذلك فإن زيادة المكررات تحسن دقة التجربة وتزيد من قوة الاختبار الإحصائي وتقلل من أطوال فترات الثقة.

(٣, ١, ٧, ٩) زيادة مجال الاستدلال الإحصائي في التجربة

تكون المكررات في بعض أنواع التجارب بفرض زيادة مجال استدلال التجربة ويكون تعريف مجتمع المعاينة أقل تقييدا والاستدلال له مجال واسع. افترض أننا نرغب على سبيل المثال في تقدير الفرق الحقيقي بين ناتج نوعين من المحاصيل في مساحة معينة، ولدينا نوعين من التربة وتلك المساحة إذا كان الفرض عمل استدلال إحصائي عن التجربة فمن البديهي أن يكون كلا من النوعين في التجربة ومن المهم أيضا أن يكون كل زوج من القطع (مكرر)، حيث يتم زرع نوعي المحصول في نوع واحد من التربة، وأن يكون منتظما على قدر الإمكان. لاحظ أنه في حالة المجتمعات العريضة الواسعة ليس من الضروري الانتظام التام داخل المكررات مثل مساحات شاسعة من الأرض. وتعاد التجربة في كثير من التجارب الحقلية خلال فترات زمنية من السنوات، وسبب الإعادة بديهي وهو أن الظروف والشروط تختلف من سنة إلى أخرى، وهذا يعطي أهمية لمعرفة تأثير السنوات على الاختلاف فيما بين المعالجات، حيث تعطى التوصيات دائما لسنوات مستقبلية. ويستخدم مواقع مختلفة لتقويم تأثير المعالجات تحت ظروف بيئية موجودة في المجتمع ذلك أنه في المساحة المراد عمل توصيات عنها يمكن اعتبار كلا من الوقت (السنوات) والموقع (المكان) أنواع من المكررات العريضة الواسعة، والفرض دائما هو زيادة مجال الاستدلال الإحصائي. تستخدم نفس المناطق بصورة متكررة في التجارب المعملية بمعنى إعادة التجربة نفسها عدة مرات، فمن الممكن بأشخاص مختلفة، ويمكن تحديد إمكانية التكرار للمعالجات تحت شروط مختلفة أن تحدث من وقت لآخر في المعمل.

(٤ , ١ , ٧ , ٩) التأثير في ضبط تباين الخطأ

تسمح المكررات بتجميع الوحدات التجريبية طبقا للاستجابة المتوقعة في حالة غياب المعالجات. يكون الغرض توزيع الاختلاف الكلي فيما بين الوحدات التجريبية بحيث يكون كبيرا بين المجموعات وصغيرا داخلها. إذا تم تخصيص المعالجات، فإن الاختلافات المشاهدة ستقيس اختلافات حقيقية بين المعالجات على أي مجموعة بطريقة أفضل من لو كنا أهملنا تجميع الوحدات. يجب أن لا يتضخم الخطأ التجريبي - المقياس لكشف أية فروق حقيقية بين المعالجات - بالاختلافات فيما بين المجموعات.

(٢ , ٧ , ٩) العوامل المؤثرة على المكررات

يتوقف عدد المكررات اللازم لتجربة ما على عدد عوامل متعددة نذكر منها:

- درجة الدقة المطلوبة.
- مقدار التباين بمواد التجربة.
- الإمكانات المتاحة للتجربة.

وسنتناول هذه العوامل بالتفصيل فيما يلي:

(١ , ٢ , ٧ , ٩) درجة الدقة المطلوبة

يجب أن يكون لدينا لأي تجربة المقدار الصحيح من الدقة وتتوقف الدقة (precision) على عدة عوامل نذكر منها:

طبيعة المعالجات وعددها، الصفات المدروسة، مقدار الاختلافات المتوقعة بين المعالجات، نوع التصميم المستخدم للصفات المدروسة ويلاحظ أن عدد المكررات يكون كبيرا كلما قل التباعد (الانحراف) المطلوب قياسه عن الفرض العدمي.

(٢ , ٢ , ٧ , ٩) مقدار التباين بمواد التجربة

من الطبيعي أن نجد بعض المواد التجريبية أكثر اختلافا وتباينا من الأخرى،

فمثلا مشكلة عدم تجانس التربة فالبعض أكثر انتظاما من الآخر. ولذا فلنفس الدقة فإننا نأخذ مكررات أقل من نوع التربة المنتظمة. أي أنه كلما ازداد مقدار التباين بمواد التجربة كلما لزم زيادة عدد المكررات. إذا تعددت الصفات المطلوب دراستها نحدد المكررات على ضوء أكثر الصفات المهمة المطلوب دراستها تباينا، حيث يلزم زيادة عدد المكررات اللازمة لدراسة هذه الصفة عما يلزم للصفات الأخرى.

(٣ , ٢ , ٧ , ٩) الإمكانات المتاحة للتجربة

يتحدد في كثير من الأحيان عدد المكررات بالامكانات المتاحة للتجربة من أدوات وأشخاص ومواد تجريبية وهذا قد تحدده التكلفة. وإذا كان أي من هذه العوامل محدودا فهذا يحدد عدد المكررات تلقائيا.

نلاحظ أن العوامل السابقة التي تؤثر على عدد المكررات تؤثر على دقة التجربة أيضا، والحقيقة أن الدقة المطلوبة تؤثر في عدد المكررات، وعدد المكررات يؤثر في الدقة أيضا. وتعتبر درجة الدقة المطلوبة أهم العوامل التي تؤثر في عدد المكررات. وأيضا لا تؤدي المكررات لتقليل الأخطاء الناتجة عن استخدام اساليب تجريبية غير مناسبة. ويجب مراعاة أن ذلك يقابله اختلافا في الناحية العملية فالأخير يجب أن تؤيده المعلومات الفنية عن الموضوع.

(٨ , ٩) التحكم في الخطأ

Error Control

يمكن تحقيق التحكم في الخطأ بالآتي:

- التصميم التجريبي.
- استخدام المشاهدات المصاحبة.
- اختيار حجم وشكل الوحدات التجريبية.

وسنتناول بالتفصيل كل واحدة فيما يلي:

(١ , ٨ , ٩) التصميم التجريبي (Experimental design)

يعتبر استخدام التصميم التجريبي في التحكم في الخطأ من الموضوعات التي تم دراستها بكثرة من بداية الربع الأول من القرن الحالي (العشرين). ولقد تناوله كثير من الباحثين أمثال كوكس، كوكرن وكوكس، كمثورن.. الخ.

يتألف التحكم في الخطأ التجريبي من تكوين تجربة بحيث إن بعض الاختلافات ما بين الوحدات التجريبية يمكن معاملتها بحيث لا تؤثر على الاختلافات بين متوسطات المعالجات. واختيار التصميم الملائم يعتمد على مصادر الاختلاف المتاحة وخبرة الباحث أيضا، حيث تجتمع الوحدات التجريبية في قطاعات كاملة (يعرف القطاع block على أنه مجموعة من الوحدات التجريبية المتجانسة) يحتوي كل قطاع على جميع المعالجات بحيث إن الاختلافات ما بين الوحدات داخل كل قطاع أصغر وفق الوحدات في القطاعات المختلفة. وتزداد دقة تلك التجربة كنتيجة للتحكم في الخطأ. وتسمى القطاعات المتشابهة مكررات والتصميم المستخدم في هذه الحالة يسمى تصميم القطاعات الكاملة تامة التعشية (completly randomized block design). يعتبر الخطأ التجريبي على الاختلافات بين الوحدات التجريبية داخل المكرر بعد تعديلها من التأثير العام للمعالجة، بمعنى أن الاختلافات بين المكررات والاختلافات ما بين المعالجات لا تدخل داخل الخطأ التجريبي.

هناك أنواع أخرى مختلفة من التصميمات كل منها يساعد على التحكم في الخطأ التجريبي وأحسن تصميم يمكن أن يستخدمه الباحث التصميم البسيط والذي يعطي الدقة المطلوبة وليس هناك أي معنى لاستخدام تصميمات معقدة لا تعطي زيادة في الدقة.

(٢ , ٨ , ٩) استخدام المشاهدات المصاحبة (Concomitant observations)

يمكن زيادة الدقة في بعض انواع التجارب باستخدام المشاهدات المصاحبة وأحد الأساليب الرياضية المسمى بتحليل التغاير (analysis of covariance). ويستخدم هذا التحليل حينما يكون جزء من الاختلاف فيما بين الوحدات التجريبية راجع

للاختلاف في بعض الصفات المقيسة أو صفات غير متحكم فيها بكفاءة مفيدة لتخصيص الوحدات التجريبية لقطاعات كاملة أو غير كاملة على أساس قوائم متشابهة.

(٩,٨,٣) حجم وشكل الوحدات التجريبية

نعطي الوحدات التجريبية الكبيرة - كقاعدة - اختلافا أقل من الوحدات الصغيرة. وتتحقق تلك القاعدة على التحديد حيثما تتوزع الأخطاء توزيعا طبيعيا بتباين عادي وكنتائج للعلاقة $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$. ومع ذلك فإن زيادة حجم الوحدة التجريبية غالبا ما تؤدي إلى تقليل عدد المكررات للتجربة وذلك لأنه في الغالب تكون المواد التجريبية محدودة. ولا يسهل الحصول على عدد ملائم من المكررات ذات قطع كبيرة عن الحصول على مكررات ذات قطع صغيرة.

في حقول التجارب الزراعية تلعب مساحة وشكل القطعة التجريبية دورا مهما في تحديد مساحة المكررة وبالتالي الدقة. فكثير من الدراسات على شكل القطع التجريبية أفادت بأنها يجب أن تكون طويلة وطبيعية للحصول على دقة عالية ويكون القطاع سواء كامل أو غير كامل مربعا تقريبا. وهذا الشكل يعطي اختلافا اكبر ما بين القطاعات وأقل من القطع. وذلك يؤدي إلى الاستفادة من استخدام القطاعات لا يوجد أثر لاختلافهم على الخطأ التجريبي ولا تؤثر أيضا على الفرق بين متوسطات المعالجات.

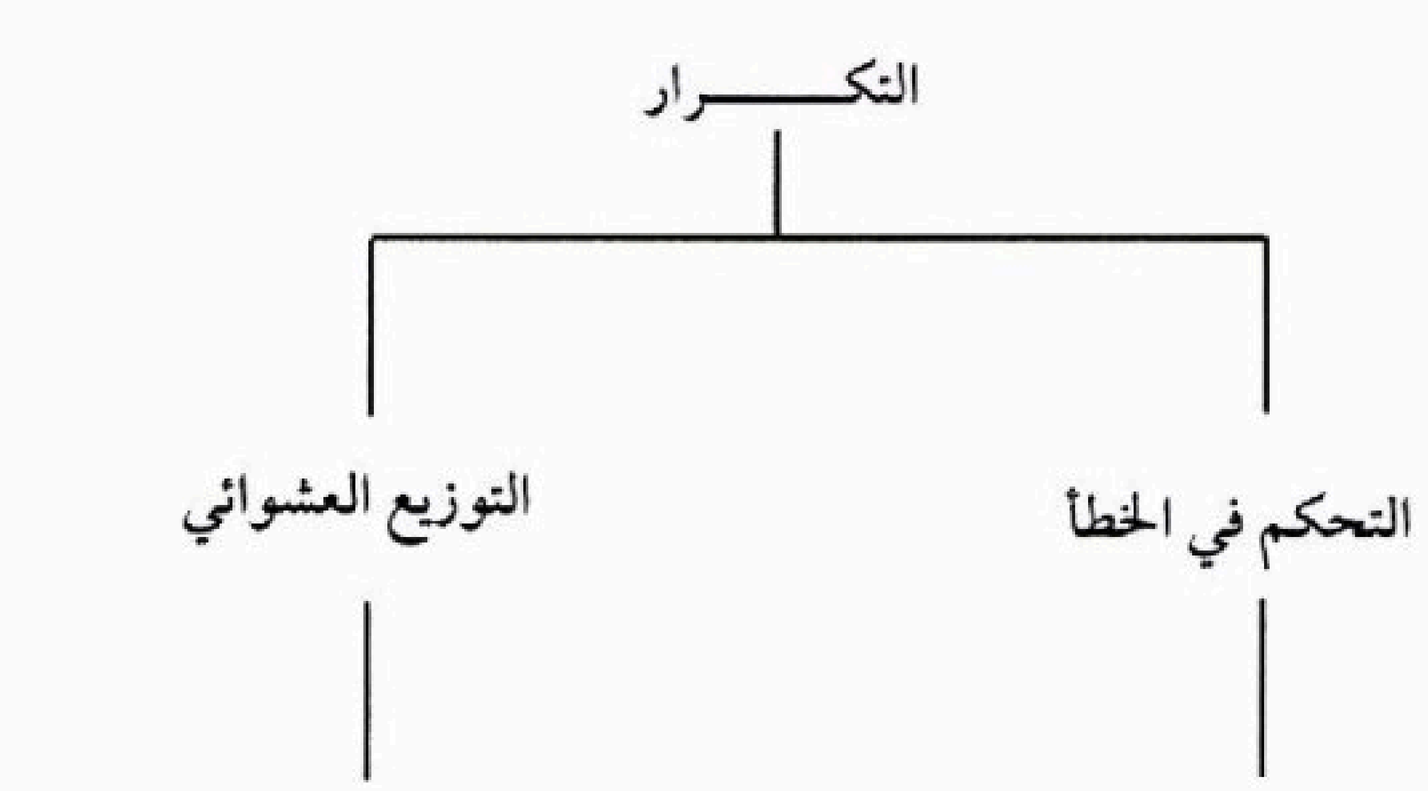
تختار الوحدات التجريبية في بعض أنواع التجربة منتظمة (متشابهة تماما) بقدر الامكان مثلا فئران التجارب تختار من نفس الولدة لتجارب التغذية. وكنتيجة فإن الخطأ التجريبي يقل ولكن في الوقت نفسه مجال الاستدلال لا يكون كبيرا.

(٩ , ٩) التعشية (العشوائية)

Randomization

يقصد بالتعشية عملية الحصول على الترتيب العشوائي بطرق موضـوعـيـة بحته، وأن تأخذ كل معالجة نفس الفرصة التي لغيرها من المعالجات، ذلك بأن نطبق هذا لقطعة تجريبية معينة ولا يكون للعامل الشخصي أي دخل في توزيع المعالجات. وتعتبر التعشية أحد ملامح تصميم التجارب كعلم حديث، ويعتبر فيشـر (Fisher) مخترعها إذا جاز التعبير. لاحظ أن التعشية وكيفما اتفق Haphazardness (الصدفة) ليس نفس الشيء والتعشية لا يمكن أن تخدم تجربة هزيلة.

ويزداد التباين في التجربة بازدياد الاختلافات غير المعلومة المصدر. ويمكن التغلب على تلك المشكلة بالتوزيع العشوائي للوحدات التجريبية. وكمـا سـبـق فالتعشية ضرورية لسلامة تقدير الخطأ التجريبي، ولتقليل التحيز في النتائج ولضمان صحة المعنوية ضد اسباب الاختلاف الذي لا يمكن التحكم فيـه. وطـرق إجـراء التوزيع العشوائي كثيرة منها سحب أوراق متطابقة في كل شيء ومرقمة بـأرقـام المعاملات منها الكيس المثالي، أو بالقاء زهرة النرد، أو استخدام الجداول العشوائية، أو الحاسب الآلي لتوليد الأرقام العشوائية، أو بغير ذلك من الطرق. ولقد رسـم فيشر العلاقة بين التكرار والتعشية والخطأ التجريبي كالتالي:



(٩, ١٠) الخطوات الواجب اتباعها لاستخدام الأسلوب الإحصائي

The Steps must Follow in Statistical Procedure

إن كل مهتم باستخدام الأسلوب الإحصائي في تصميم وتحليل التجارب عليه أن يتبع الخطوات المقترحة التالية:

(٩, ١٠, ١) صياغة المشكلة والفروض

تبدأ الدراسة العلمية بصياغة المشكلة على صورة سؤال أو أسئلة يريد الباحث الإجابة عليها أو فروض يريد اختبارها. وينبغي أن تكون الأسئلة واضحة الصياغة دقيقة، وذلك تزيد كفاءة التجربة، ويمهد السبيل نحو التقدم بسرعة لتحليل المشكلة تحليلا عمليا.

(٩, ١٠, ٢) تحليل المشكلة

يساعد في إعادة صياغة المشكلة، والفروض، ووصف المجتمع الذي تجري عليه الدراسة، ومعرفة ووضوح الأسباب التي من أجلها أجرى البحث، والنظر في المراجع المرتبطة بالمشكلة.

(٩, ١٠, ٣) اختيار طريقة الدراسة (البحث)

تختلف الطرق البحثية طبقا لظروف كل بحث وتتوقف الطريقة المناسبة على حالة الدراسة ومجالها وعليه تتضمن الأفكار الرئيسة التالية:

- اختيار المعالجات والعوامل ومستوياتها إذا كانت متعددة العوامل.
- اختيار الصفات المراد اختبارها وقياسها.
- اختيار تصميم التجربة المناسبة وتحديد عدد المكررات وتحديد حجم العينة وطريقة أخذها.
- التحكم في تأثير الوحدات التجريبية المتجاورة بعضها على بعض.
- الأشكال العامة للجداول والبيانات المتوقعة.

- الأشكال العامة للتحليلات المطلوب إجراؤها.
- حساب تكلفة البحث والمواد والأدوات والأشخاص اللازمة للتجربة ويحدد الباحث على ضوء البنود السابقة إمكانياته وقد يلجأ إلى إعادة صياغة الفرض، أو تغيير التجربة حسب ما يتراءى له على ضوء الدراسة.

(٤ , ١٠ , ٩) إجراء التجربة

هذه مرحلة جمع البيانات الحقيقية، ويجب أن يراعي المجرب أنها تسير بناء على خطة البحث وهدفه، ويجب مراعاة التعشية ودقة القياس وانتظام وتماثل ظروف التجربة.

(٥ , ١٠ , ٩) تحليل البيانات

ينبغي تحليل البيانات في ضوء الظروف التجريبية تحليلا كاملا، وينبغي تفسير النتائج في ضوء الأدلة الإحصائية، وظهرت في الأعوام الحالية حزم برامج جاهزة إحصائية ساعدت على سرعة إجراء تحليل البيانات وبدقة عالية، كما تساعد على معرفة ملائمة النماذج الإحصائية ورسمها، وأيضا ملائمة الفروض الإحصائية التي كانت قاعدة للتحليل الإحصائي.

(٦ , ١٠ , ٩) إعداد التقرير والنتائج والتوصيات

ينبغي كتابة تقرير واضح عن التجربة وبعد انتهاء التحليل، واستخلاص النتائج الإحصائية على ضوء الواقع العملي والعلمي لموضوع البحث.

(١١ , ٩) محاذير عامة

General Warning

إن استخدام الطرق الإحصائية يزيد من كفاءة استخدام التجارب في جميع مجالات العلوم وتعطي مذاقا خاصا للنتائج. ولكي تؤدي الأساليب الإحصائية

إلى المرجو منها يجب أن يكون في ذهن الباحث دائما الموضوعات التالية:

- استخدام المعلومات غير الإحصائية عن المشكلة.
- بساطة تصميم التجربة وبساطة التحليل المستخدم بقدر الإمكان
- مراعاة والتعرف على الفرق بين المعنوية الإحصائية والمعنوية المعملية.
- إن التجارب عادة يمكن تكرارها.

(٩, ١٢) أنواع تصميمات التجارب

Kinds of Experimental Designs

تتعدد تصميمات التجارب ويلائم كل منها هدفا معينا، ولو أحسن الباحث اختيار التصميم المناسب ارتفعت قيمة النتائج المتحصل عليها، ويمكن تقسيم التجارب إلى أقسام طبقا لعدد عوامل الدراسة أو حالة القطاع.

(٩, ١٢, ١) التقسيم طبقا لعدد عوامل الدراسة

- تصميمات بسيطة وهي التي تتضمن عامل واحد فقط للدراسة مع ثبات العوامل الأخرى.
- تصميمات التجارب العاملية وهي تلك التي تتضمن أكثر من عامل من عوامل الدراسة.

(٩, ١٢, ٢) التقسيم طبقا لحالة القطاع

- تصميمات القطاعات الكاملة من أهمها:

التصميم تام التعشية - تصميم القطاعات الكاملة العشوائية - تصميم المربع اللاتيني.

- تصميمات القطاعات الناقصة ومنها:

تصميم القطع المنشقة والشبكية.

(١٣, ٩) تمارين

Exercises

١ - اذكر ما تعرفه عن:

التجريب والتجربة - أهداف التجربة.

٢ - اذكر ما تعرفه عن:

الوحدة التجريبية والمعالجة - الخطأ التجريبي.

٣ - اذكر ما تعرفه عن:

المكررات ووظائفها - التحكم في الخطأ - التعشية - أنواع تصميمات التجارب.

٤ - ماهي الخطوات الواجب اتباعها لاستخدام الإحصائي في تصميم وتحليل التجارب؟

٥ - ما هي المحاذير العامة التي يجب أن تكون في ذهن الباحث عند استخدام الأساليب الإحصائية في تصميم وتحليل التجارب.

الفصل العاشر

تحليل التباين في اتجاه واحد
One Way Analysis of Variance

● مقدمة ● تحليل التباين لأي عدد من المجموعات بمكررات متساوية ● النموذج الرياضي والعلاقات الرياضية ● نموذج التأثيرات الثابتة ● تحليل التباين في حالة عدم تساوي المكررات (الحالة غير المتزنة) ● نموذج التأثيرات العشوائية ● تصميم التام التعشية ● تمارين

(١٠ , ١) مقدمة
Introduction

ناقشنا في الفصول السابقة الطرق المختلفة لمقارنة متوسطي عينتين (التجربة التي تشتمل على معالجتين فقط)، ولكن في كثير من التطبيقات العملية نجد أن التجارب تشتمل على أكثر من معالجتين، حيث يكون لدينا أكثر من متوسطين للمقارنة.

افترض على سبيل المثال: أجرى أحد الباحثين تجربة على نبات القمح لدراسة تأثير نترات الكالسيوم (كسماد) على كمية محصول الحبوب. نتج لكل مستوى من نترات الكالسيوم كمية محصول معين للحبوب. اهتم الباحث باختبار تساوي متوسطات كميات المحاصيل للحبوب للمستويات المختلفة. فإذا كان لدينا عدد a مستوى من السماد (نترات الكالسيوم) بحيث $a \geq 2$، ونريد استخدام اختبار t لكل ازواج من المتوسطات، فهذا يعني أن لدينا$\binom{a}{2}$ مقارنة. فمثلا إذا كانت

a=6 فإنه يكون لدينا $15 = \binom{6}{2}$ مقارنة وبالتالي نجري 15 اختبارا لاختبار t. إذا فكرنا في نظرية الاستدلال الإحصائي وخاصة الخطأ من النوع الأول α عند مستوى معنوية وليكن $\alpha = 0.05$. نجد أن احتمال قبول فرض العدم وهو صحيح لكل اختبار هو $(1-\alpha) = 0.95$، وبذلك يصبح احتمال قبول فرض العدم (H_0) وهو صحيح للاختبارات الخمسة عشر على افتراض أنها مستقلة هو $(0.95)^{15} = 0.4633$ ، وبذلك يصبح الخطأ من النوع الأول في هذه الحالة كبيرا جدا $\alpha = 1 - 0.4633 = 0.5367$. وبالتالي تصبح الحاجة شديدة إلى أسلوب آخر وهذا الأسلوب هو أسلوب تحليل التباين والواقع أن له تطبيقات أكثر مما ذكر في المثال السابق.

سنناقش في هذا الفصل والذي يليه الطرق الخاصة بمقارنة المعالجات وخاصة حينما تكون موزعة عشوائيا على الوحدات التجريبية. وعلى الرغم من أنه يمكن استخدام تحليل التباين في حالة عينتين مستقلتين ويعطي نفس النتائج فيما لو استخدمنا اختبار t، إلا أنه أكثر فائدة حينما يكون هناك أكثر من عينتين في التجربة. ويعتبر السير رونالد فيشر (Ronald A. Fisher 1890 -1962) أول من قدم تحليل التباين كأسلوب لتحليل بيانات التجارب المختلفة. كما يعتبر تحليل التباين الآن من أوسع طرق تحليل البيانات في المجالات البحثية سواء كانت صناعية أو زراعية أو أحيائية (بيولوجية) أو أي مجال آخر يكون فيه البيانات مقيسة كميا. كما أنه الأسلوب الأكثر أهمية في الاستدلال الإحصائي.

ويعرف تحليل التباين على أنه عملية حسابية يتم فيها تجزيء مجموع المربعات الكلي إلى مكونات متعلقة بمصادر الاختلاف، أي يتم تجزيء التباين الكلي إلى مكوناته المعروفة (مصدرا للاختلاف أو مصدرين للاختلاف. . .) إلى الجزء الذي يرجع إلى أسباب غير معلومة والذي يسمى الخطأ التجريبي.

(٢ , ١٠) تحليل التباين لأي عدد من المجموعات بمكررات متساوية

Analysis of Variance for any Groups with Equal Replects

إذا كان لدنيا عدد المجموعات (المعالجات) a وكل معالجة مكررة n مرة ففي هذه الحالة يصبح التباين بمكررات متساوية، ويسمى تحليل التباين المتزن. **والمفاهيم الأساسية لتحليل التباين نلخصها كما يلي:**

- إنه يمكن التفرقة بين مصدرين للاختلاف المصدر الأول المعروف بين المجموعات (المعالجات) والمصدر الثاني والمعروف داخل المجموعات (الخطأ). يعكس التباين بين المجموعات حجم الاختلاف بين متوسطات المجموعات (المعالجات)، وكلما زاد الفرق بين تلك المتوسطات كلما زاد التباين. أما المصدر الثاني داخل المجموعات (الخطأ) يعكس تشتت واختلاف القيم داخل كل مجموعة (معالجة). يشار إلى مجموع التشتت داخل جميع المعالجات بالخطأ التجريبي.

- يكون الفرض العدمي الإحصائي(H_0) هو تساوي متوسطات المعالجات فمثلا $H_0 : \mu_1 = \mu_2 =...= \mu_a$.

- نحسب قيمة F_0 بقسمة متوسط مربعات بين المجموعات (مجموع مربعات بين المجموعات (المعالجات) / درجات حريتها) على متوسط مربعات داخل المجموعات (مجموع مربعات داخل المجموعات (الخطأ) / درجات حريتها). بمقارنة قيمة F_0 المحسوبة بقيمة $F_{\upsilon_1, \upsilon_2, \alpha}$ الجدولية عند درجات حرية معينة(υ_1, υ_2) يمكن رفض أو قبول فرض العدمي H_0.

سنشرح في المثال التالي طريقة حساب مجاميع المربعات وكيفية إجراء اختبار F.

مثال (١ , ١٠)

أُجريت دراسة لمعرفة الفرق بين تأثير ثلاث طرق مختلفة لتدريس مبادىء الحساب لطلاب المرحلة الأولى الابتدائية. أختير 27 تلميذا عشوائيا. تم تخصيص 9 تلاميذ منهم بطريقة عشوائية لكل طريقة من الطرق المختلفة السابقة. تم اختبار جميع التلاميذ بعد فترة معينة وكانت نتائج الاختبارات كما في جدول (١ , ١٠) التالي (بيانات افتراضية)

جدول (١ , ١٠)

رقم الطالب	1	2	3	4	5	6	7	8	9	المجموع
الطريقة الأولى	4	5	4	3	6	10	1	8	5	46
الطريقة الثانية	12	8	10	5	7	9	14	9	4	78
الطريقة الثالثة	1	3	4	6	8	5	3	2	2	34

المطلوب معرفة هل هناك فرق معنوي بين طرق التدريس المختلفة عند مستوى معنوية $\alpha = 0.05$.

الحل

لمعرفة هل هناك فرق معنوي بين طرق التدريس المختلفة عند مستوى معنوية $\alpha = 0.05$ نقوم بعمل الخطوات التالية:

أولا: اعتبر المشاهدة y_{ij} هي المشاهدة رقم j حيث $j = 1,2,3,...,9$ للمعالجة رقم i حيث $i = 1,2,3$ في مثالنا هذا (لاحظ أن $a = 3, n = 9$) ثم نقوم بتربيع كل قيمة أي أن y_{ij}^2 وننظم البيانات كما في جدول (٢ , ١٠) كما يلي:

جدول (٢ , ١٠)

	الطريقة الاولي		الطريقة الثانية		الطريقة الثالثة		
	y_{1j}	y_{1j}^2	y_{2j}	y_{2j}^2	y_{3j}	y_{3j}^2	
	4	16	12	144	1	1	
	5	25	8	64	3	9	
	4	16	10	100	4	16	
	3	9	5	25	6	36	
	6	36	7	49	8	64	
	10	100	9	81	5	25	
	1	1	14	196	3	9	
	8	64	9	81	2	4	
	5	25	4	16	2	4	
[1] $Y_{i.}$	46		78		34		$Y_{..}=158$
[2] Σy_{ij}^2		292		756		168	$\Sigma\Sigma y_{ij}^2=1216$
[3] $(\sum Y_{i.})^2/9$	235.11		676		128.44		$\Sigma(Y_{i.})^2/9=1039.55$
[4] $\overline{Y}_{i.}$	5.111		8.6667		3.7778		$\overline{Y}_{..}=5.852$

ثانيا: نلاحظ أن $Y_{i.}$ وهو مجموع المشاهدات للعمود الأول داخل كل طريقة $\sum_j y_{ij} = Y_{i.}$ ، $\sum_j y_{ij}^2$ مجموع العمود الثاني داخل كل طريقة كما هو موضح في جدول (٢ , ١٠) السابق.

ثالثا: نحسب $158 = \sum_{i=1}^{a} Y_{i.} = Y_{..}$ وهو مجموع الصف [1] في الجدول (٢ , ١٠) ثم نحسب مجموع الصف [2] في جدول (٢ , ١٠) وهو $1216 = \sum y_{i,j}^2 = \sum(\sum y_{ij}^2)$.

رابعا: نحسب $Y_{i.}^2/9$ وذلك بتربيع كل قيمة في الصف رقم [1] في جدول (٢ , ١٠) وقسمتها على 9 عدد المكررات في كل طريقة ثم جمع الناتج لنحصل على $\Sigma(Y_{i.})^2/g = 1039.55$ كما هو موضح في الصف رقم [3] في جدول (٢ , ١٠).

خامسا: ممكن حساب $\overline{Y}.. = \frac{Y..}{27} = 5.852$ وحساب كل من $\overline{Y}i. = \frac{Y_{i.}}{9}$ في الصف [4] في جدول (٢ , ١٠) حيث ممكن الاستفادة بها فيما بعد.

سادسا: من الحسابات السابقة الملخصة في جدول (٢ , ١٠) يمكن حساب الآتي:
معامل التصحيح CF حيث

$$CF = \frac{Y_{..}^2}{n\,a} = \frac{Y_{..}^2}{27} = \frac{(\sum_{i,j} y_{ij})^2}{27} = \frac{(158)^2}{27} = 924.593$$

مجموع المربعات الكلي $= \sum_{i,j}(y_{ij} - \overline{Y}..)^2$ ويرمز له بالرمز SS_{Tot} أي أن

$$SS_{Tot} = \sum_{i,j} y_{ij}^2 - CF = 1216 - 924.593 = 291.407$$

ومجموع المربعات الكلي له درجات حرية هي $n\,a - 1 = 27 - 1 = 26$ درجة

مجموع مربعات بين المجموعات (مجموع مربعات المعالجات)

$= n\sum_{i=1}^{a}(\overline{Y}_{i.} - \overline{Y}..)^2$ ويرمز له SS_{Trt} حيث إن

$$SS_{Trt} = \sum_{i=1}^{a} \frac{(Y_{i.})^2}{n} - CF = 1039.55 - 924.593 = 114.957$$

وله درجات حرية مصاحبة له هي $a - 1 = 3 - 1 = 2$ درجة.

مجموع مربعات داخل المجموعات (الخطأ) $\Sigma(y_{ij} - \overline{Y}_{i.})^2$ ويرمز له بالرمز SSE حيث إن

$$SSE = SST_{Tot} - SST_{Trt}$$
$$= 291.407 - 114.957 = 176.450$$

ودرجات الحرية المصاحبة لمجموع مربعات الخطأ SSE تحسب بالطرح فتكون $26 - 2 = 24$ درجة.

ومن ذلك يمكن حساب متوسط مجموع المربعات وذلك بقيمة كل مجموع مربعات على درجات الحرية المصاحبة له أي أن

$$MST_{Trt} = \frac{SST_{Trt}}{a - 1} = \frac{114.957}{2} = 57.478$$

$$MSE = \frac{SSE}{na - a} = \frac{176.4501}{27 - 3} = \frac{175.4501}{24} = 7.35$$

وتصبح قيمة F_0 المحسوبة لاختبار أن متوسطات الطرق الثلاث متساوية أي أن الفرض العدمي (H_0) هو:

$$H_0 = \mu_1 = \mu_2 = \mu_3$$

ضد الفرض البديل H_1 أي أن

H_1: واحدة على الأقل لا تساوي

أي أن

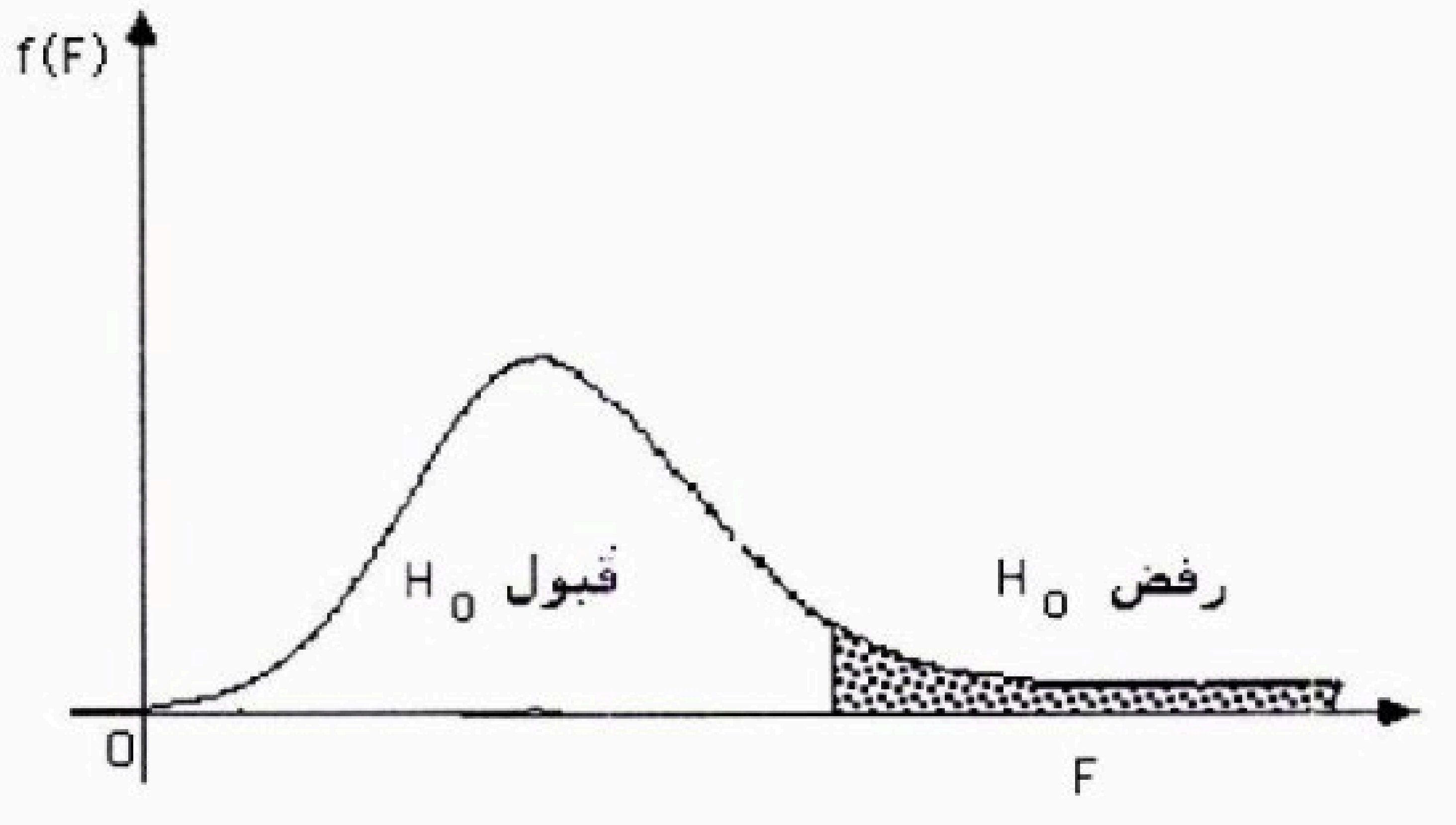
$$F_0 = \frac{57.478}{7.35} = 7.82$$

ويكون لدينا جدول تحليل التباين (٣ , ١٠) ويسمى جدول أنوفا ANOVA كما يلي:

جدول (٣ , ١٠)

المصدر	مجموع المربعات SS	درجات الحرية d.f.	متوسط مجموع المربعات MSS	F_0
بين المجموعات	114.957	2	57.478	7.82
داخل المجموعات	176.4501	24	7.35	
الكلي	291.4071	26		

ونقارن هذه القيمة $F_0 = 7.82$ المحسوبة في جدول (٣ , ١٠) بقيمة $F_{2, 24, 0.05}$ الجدولية من جدول F المعطى في آخر الكتاب لدرجات حرية (2, 24) ومستوى معنوية $\alpha = 0.05$ حيث نجد أن $F_{2, 24, 0.05} = 3.40$ ونوضح منطقة قبول H_0 بالفترة (0, 3.40)، ومنطقة رفض H_0 بالجزء المظلل في الرسم شكل (١ , ١٠) كما يلي:



شكل (١ , ١٠)

نلاحظ أن قيمـة $F_0 = 7.82$ المحسوبة في جدول تحليل التبـايـن (٣, ١٠) واقعة في منطقة رفض H_0 أي أننا نرفض فرض العدم H_0. أي أن هناك فرقا معنويا بين متوسطات الطرق الثلاث المختلفة للتدريس.

ملاحظة

نلاحظ أن متوسط مربعات الخطأ يعرف بالمـقـدار S^2 حيث يؤخذ كتقدير للتبايـن σ^2 في حالة ما يكون النموذج ثابتا وله تبايـن σ^2. ويمكن تقدير الخطـأ المعياري لكل متوسط وللفرق بين متوسطين كما يلي:

$$S_{\overline{Y}_{i.}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{7.35}{9}} = 0.9036$$

$$S_{\overline{Y}_{i.} - \overline{Y}_{i'.}} = \sqrt{\frac{2S^2}{n}} = \sqrt{2\left(\frac{7.35}{9}\right)} = 1.278$$

حيث $i \neq i'$

وهذه الإحصاءات مفيدة عند إجراء القياسات المتعددة كما سنرى في الفصل القادم.

(٣, ١٠) النموذج الرياضي والعلاقات الرياضية

Mathematical Model

افترض أن لدينا مجموعة من المعالجات أو عامل وحيد له عدة مستويـات يراد المقارنة بينهم. فإن المشاهدات الناتجة y_{ij} ; $i = 1,2,..,a$; $j = 1,2,...,n$ كما ستوضح في جدول (٤, ١٠) بيانات تحليل التباين ذي اتجاه واحد كما يلي:

جدول (٤ , ١٠)

		المشاهدات			
المعالجات	1	y_{11}	y_{12}	...	y_{1n}
	2	y_{21}	y_{22}	...	y_{2n}
		:	:	:	:
		:	:	:	:
		:	:	:	:
	a	y_{a1}	y_{a2}	...	y_{an}

يمكن وصف المشاهدات y_{ij} بالنموذج الخطي التالي:

$$Y_{ij} = \mu + \tau_i + \in_{ij} \qquad (١٠ , ١)$$

حيث إن: $j = 1,2,.., n$, $i = 1,2,..., a$

y_{ij} المشاهدة رقم i, j، μ متوسط عام لجميع المعالجات (معلمة ثابتة)

τ_i تأثير المعالجة رقم i، $\in_{ij}$ الخطأ العشوائي

ويكون الفرض من النموذج هو تقدير تلك المعالم وإجراء الاختبارات الإحصائية المختلفة. ويفترض في هذا النموذج أن الأخطاء العشوائية لها متوسط صفر$(E(\in_{ij}) = 0)$ وتباين σ^2 $(V(\in_{ij}) = \sigma^2)$. وأن σ^2 ثابتة لجميع المعالجات أو لجميع مستويات العامل. يسمى النموذج تحليل التباين ذي الاتجاه الواحد وذلك لأن هناك عاملا واحدا فقط تحت الدراسة.

يتطلب النموذج أيضا أن تتشابه جميع الوحدات التجريبية ويسمى تصميم التجربة الذي ينطبق عليه تلك الفروض تصميم التام التعشية (completely randomized

design). لاحظ أن ذلك النموذج (١ , ١٠) قد يصف حالتين مختلفتين بالنسبة للمعالجات: **أولهما** أن المعالجات كلها (a) مختارة من المجرب لاختبارها وفي هذه الحالة تطبق النتائج فقط على تلك المعالجات ولا يمكن تعميمها على معالجات أخرى لم تستخدم في التجربة، وقد نهتم بتقدير التأثيرات لتلك المعالجات(μ, τ, σ^2) وذلك يسمى نموذج التأثيرات الثابتة. **والثاني** أن تكون المعالجات، فيه عشوائية من مجتمع أكبر من المعالجات وبذلك فإننا نريد أن نعمم النتائج على ذلك المجتمع، وتعتبر τ_i في هذه الحالة متغيرا عشوائيا ونريد معرفة معلومات حول تباينها، وتقدير ذلك التباين، وهذا ما يسمى بنموذج التأثيرات العشوائية. وسنتناول بالشرح والأمثلة كل من النموذجين التأثيرات الثابتةوالتأثيرات العشوائية كما يلي.

(٤ , ١٠) نموذج التأثيرات الثابتة

Model of Fixed Effects

تعرف التأثيرات الثابتة للمعالجات τ_i في ذلك النموذج على أنها انحرافات من المتوسط العام بحيث إن

$$\sum_{i=0}^{a} \tau_i = 0 \qquad (١٠ , ٢)$$

ويمكن التعبير بالرموز الآن عن المتوسطات كما يلي:

$$Y_{i.} = \sum_{j=1}^{n} y_{ij} \quad , \qquad \overline{Y}_{i.} = \frac{Y_{i.}}{n} \;\; ; \; i = 1,2,\ldots, a$$

$$Y_{..} = \Sigma\Sigma\, y_{ij} \quad , \qquad \overline{Y} = \frac{Y_{..}}{N} \quad , \qquad N = n\,a$$

$$E(y_{ij}) = \mu + \tau_i = \mu_i$$

ويصبح اختبار تساوي متوسطات المعالجات كما يلي:

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_a$$

على الأقل واحدة i,j لا تساوي $H_1: \mu_i \neq \mu_j$

وحيث إن جميع المتوسطات تشتمل علـى μ كمقدر ثابت فيصبح ذلـك الاختبار نفسه معبرا عنه بالتأثيرات الثابتة بالمعالجات τ_i كما يلي:

$$H_0: \tau_1 = \tau_2 = \ldots = \tau_a = 0$$

على الأقل واحدة لا تساوي $H_1: \tau_i \neq 0$

ذلك فإنه يمكننا استخدام تساوي المتوسطات أو تأثيرات المعالجات التي تـسـاوي صفرا كأختبار فروض حول المعالجات المدروسة.

(١ , ٤ , ١٠) مجاميع المربعات

كما سبق أن ذكرنا فإن كلمة تحليل التباين أحدث من تقسيم التباين (الاختلاف) الكلي إلى مكوناته وأجزائه المختلفة ومن المعروف أن مجموع المربعات الـكـلـي الصحيح هو:

$$SS_{Tot} = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \overline{Y}_{..})^2$$

يستخدم كمقياس للتباين الكلي للبيانات ومن البديهي أن درجات الحرية المصاحبة تصبح ($n - 1 = N - 1 = \upsilon$ درجة). وناتج قسمة $\frac{SS_{Tot}}{N-1}$ يعطي تباين العينة كلهـا ويمكن كتابة SS_{Tot} كما يلي:

$$\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \overline{Y}_{..})^2 = \sum_{i=1}^{a} \sum_{j=1}^{n} \{(\overline{Y}_{i.} - \overline{Y}_{..}) + (y_{ij} - \overline{Y}_{i.})\}^2$$

وذلك يتم بطرح واضافة $\overline{Y}_{i.}$ فإن

$$\sum\sum\left(y_{ij}-\overline{Y}_{..}\right)^2 = n\sum_{i=1}^{a}\left(\overline{Y}_i-\overline{Y}_{..}\right)^2+\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-Y_i\right)^2$$

$$+2\sum_{i=1}^{a}\sum_{j=1}^{n}(\overline{Y}_{i.}-\overline{Y}_{..})(y_{ij}-\overline{Y}_{i.}) \qquad (١٠,٣)$$

ولكن

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(\overline{Y}_{i.}-\overline{Y}_{..})(y_{ij}-\overline{Y}_{i.})=0$$

لأن

$$\sum_{j=1}^{n}(y_{ij}-\overline{Y}_{i.})=Y_{i.}-n\overline{Y}_{i.}=0$$

وبذلك تصبح العلاقة (٣, ١٠) السابقة على الشكل التالي:

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\overline{Y}_{..})^2=n\sum_{i=1}^{a}(\overline{Y}_{i.}-\overline{Y}_{..})^2+\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\overline{Y}_{i.})^2$$

(١٠, ٤)

نلاحظ من العلاقة (٤, ١٠) السابقة أن التباين الكلي مقاسا بمجموع المربعات الكلي الصحيح تم تقسيمه إلى جزئين: **الأول** مجموع مربعات الفروق بين المتوسطات والمتوسط العام. **والثاني** مجموع مربعات الاختلافات بين المشاهدات داخل كل معالجة ومتوسط المعالجة. ويعتبر الحد الأول في الطرف الأيمن للعلاقة (٤, ١٠) مقياس للاختلافات بين المعالجات والحد الثاني في نفس الطرف يعتبر راجعا للخطأ العشوائي. أي يمكن كتابة مجموع المربعات الكلي في (٤, ١٠) بالصورة التالية:

$$SS_{Tot} = SS_{trt} + SSE \qquad (١٠,٥)$$

حيث إن:

SS_{trt} يسمى مجموع مربعات المعالجات (بين المعالجات)،

SSE يسمى مجموع مربعات الخطأ.

كما نعلم أن لدينا درجات حرية $n\,a - 1 = N - 1$ مصاحبة للكلي(SS_{Tot}). وأنه لدينا درجات حرية $a - 1$ مصاحبة (SS_{trt}) وداخل كل معالجة لدينا n مكرر أي لدينا $n - 1$ درجة حرية. لتقدير الخطأ التجريبي لكل معالجة ولدينا a معالجة فإنه يصبح لدينا $a(n - 1) = n\,a - a = N - a$ درجة حرية مصاحبة للخطأ ونلاحظ أن

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \overline{Y}_{i.})^2 = \sum_{i=1}^{a}[\sum_{j=1}^{n}(y_{ij} - \overline{Y}_{i.})^2]$$

$$= \sum_{i=1}^{a}[(n-1)\,S_i^2]$$

حيث إن S_i^2 هو تباين العينة للمعالجة رقم i ويصبح المقدار

$$\frac{\sum_{i=1}^{a}(n-1)\,S_i^2}{\sum_{i=1}^{a}(n-1)} = \frac{SSE}{N - a}$$

هذا يعني أن SSE / N-a تقدير للتباين داخل كل معالجة. البعض قد يستخدم SS_{trt} لتقدير التباين أيضا وذلك في حالة تساوي متوسطات المعالجات، المقداران

$$MSE = \frac{SSE}{N - a} \quad , \quad MSS_{trt} = \frac{SS_{trt}}{a - 1}$$

يطلق عليهما متوسط المربعات ويمكن حساب توقعهما طبقا للنموذج الثابت وتكون النتيجة كما يلي:

$$E(MSS_{trt}) = \sigma^2 + \frac{n\Sigma\tau_i^2}{\alpha - 1}$$

$$E(MSE) = \sigma^2$$

ومن ذلك يبدو واضحا أن لاختبار تساوي متوسطات المعالجات يمكن تكوينه من مقارنة MSS_{trt} و MSE كما سنرى فيما يلي:

(٢ , ٤ , ١٠) التحليل الإحصائي

لإجراء الاختبار الإحصائي أن متوسطات المعالجات متساوية يمكن استخدام فرض العدم

$$H_0 : \mu_1 = \mu_2 = ... = \mu_a$$

أو

$$H_0 : \tau_1 = \tau_2 = ... = \tau_a = 0$$

ويجب أن نفرض أيضا أن الاخطـاء $\in_i$ تتوزع توزيعا طبيعـيـا $\in_i \sim N(0, \sigma^2)$ وأنها مستقلة وباستخدام بعض النظريات الإحصائية نقـول إن SS_{trt}/σ^2 تتوزع بمربع كاي (χ^2_{a-1}) بدرجات حرية a - 1 وأن $\frac{SSE}{\sigma^2}$ أيضا تتوزع مربع كاي (χ^2_{N-a}) بدرجات حرية N - a وأنهما مستقلان إحصائيا، ومن ذلك ينتج أنه وتحت فرض العدم (H_0) أن النسبة:

$$F_0 = \frac{SS_{trt}/(a - 1)}{SSE / (N - a)} = \frac{MSS_{trt}}{MSE}$$

حيث F_0 تتوزع طبقا لتوزيع F بدرجات حرية [(a-1), (N-a)] وأن F_0 تصبح إحصاءة لاختبار أن: متوسطات المعالجات متساوية ونرفض H_0 إذا كان $F_0 > F_{a-1, N-a, \alpha}$ حيث F_0 هي المحسوبة من ناتج قسمة متوسطات مجموع مربعات المعالجات ومجموع مربعات الخطأ وأن $F_{a-1,N-a,\alpha}$ قيمة من الجدول (٤) آخر الكتاب لتوزيع F عند مستوى معنوية α وبدرجات حرية [(a−1), (N−a)]. وتعتبر المعادلات الآتية من صيغ حسابية لمجاميع المربعات المختلفة

$$SS_{Tot} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{Y_{..}^2}{N}$$

$$SS_{trt} = \Sigma \frac{Y_{i.}^2}{n} - \frac{Y_{..}^2}{N}$$

$$SSE = SS_{Tot} - SS_{trt}$$

ونلخص المعلومات السابقة في جدول (٥ , ١٠) التالي ويطلق عليه جـدول تحليـل التبـاين (ANOVA) وهي اختصار (Analysis of variance) .

جدول (٥ , ١٠)

مصدر الاختلاف	مجموع المربعات SS	درجات حرية d.f.	متوسط مجموع المربعات MSS	توزيع F
بين المعالجات	SS_{trt}	a -1	MSS_{trt}	$F_0 = \frac{MSS_{trt}}{MSE}$
الخطأ داخل المعالجات	SSE	N - a	MSE	
الكلي	SS_{Tot}			

(٥ , ١٠) تحليل التباين في حالة عدم تساوي المكررات (الحالة غير المتزنة)
Analysis of Variance for Groups with Unequal Replects

تختلف عدد المشاهدات داخل كل معالجة في بعض التجارب أحادية العامل، وفي هذه الحالة يسمى التصميم غير متزن، ولحسن الحظ فإن تحليل التباين السابق لا زال من الممكن استخدامه لتلك النوعية من التجارب وبنفس الفروض، ولكن يتم تعديل بسيط وحساب مجاميع المربعات لأنه سيصبح لدينا n_i مشاهدة للمعالجة i وتصبح $N = \sum_{i=1}^{a} n_i$ وتصبح الصيغ الحسابية كما يلي:

$$SS_{Tot} = \sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{Y_{..}^2}{N}$$

$$SS_{trt} = \sum_{i=1}^{a} \frac{Y_{i.}^2}{n_i} - \frac{Y_{..}^2}{N}$$

لاحظ أن الخطأ المعياري للمتوسط $S_{\overline{X}_i}$ هو

$$S_{\overline{Y}_i} = \sqrt{\frac{S^2}{n_i}}$$

ويكون الخطأ المعياري بين متوسطين $S_{\overline{Y}_i - \overline{Y}_i'}$ حيث $i \neq i'$ هو

$$S_{\overline{Y}_i - \overline{Y}_i'} = \sqrt{S^2\left(\frac{1}{n_i} + \frac{1}{n_i'}\right)}$$

يجب أن ننوه هنا أنه يجب اختيار - على قدر الإمكان - تصميما متزنا، لأنه في حالة تساوي المكررات فإن الاختبار الإحصائي يصبح أكثر قوة وأيضا أقل حساسية للبعد عن فرضية تساوي التباين للمعالجات.

مثال (٢ , ١٠)

افترض في مثال طرق التدريس (١ , ١٠) أن عدد التلاميذ 25 تلميذا فقط، ثم تم تخصيص 7 تلاميذ للطريقة الأولى و 9 تلاميذ لكل من الطريقة الثانية والثالثة، وأن النتائج كما هي فيما عدا النتيجتين 5,8 درجات في نهاية الطريقة الأولى جدول (٢ , ١٠). هل يوجد فرق معنوي بين طرق التدريس المختلفة عند مستوى معنوية $\alpha = 0.01$.

الحل

لا يختلف الحل كثيرا عما سبق شرحه في مثال (١ , ١٠) السابق ونلخص نتائج الحل في جدول (٦ , ١٠) كما يلي:

جدول (٦ , ١٠)

			الطريقة الأولى		الطريقة الثانية		الطريقة الثالثة
	y_{1j}	y_{1j}^2	y_{2j}	y_{2j}^2	y_{3j}	y_{3j}^2	
	4	16	12	144	1	1	
	5	25	8	64	3	9	
	4	16	10	100	4	16	
	3	9	5	25	6	36	
	6	36	7	49	8	64	
	10	100	9	81	5	25	
	1	1	14	196	3	9	
	-	--	9	81	2	4	
	-	--	4	16	2	4	
[1] $Y_{i.}$	33		78		34		$Y_{..}=145$
[2] Σy_{ij}^2		203		756		168	$\Sigma\Sigma y_{ij}^2=1127$
[3] $(Y_i)^2/n_i$	155.57		676		128.44		$\Sigma(Y_{i.})^2=960.01$
[4] $\overline{Y}_{i.}$	4.71		8.67		3.78		$\overline{Y}_{..} = 5.8$

ويصبح معامل التصحيح CF هو

$$CF = \frac{(145)^2}{25} = 841$$

وكذلك مجموع المربعات الكلي SS_{Tot} هو:

$$SS_{Tot} = 16 + 25 + ... + 4 - 841$$
$$= 1127 - 841 = 286$$

وأيضا مجموع المربعات بين المعالجات SS_{trt} هو:

$$SS_{trt} = 155.57 + 767 + 128.44 - 841 = 119.01$$

ويصبح جدول تحليل التباين (٧ , ١٠) كما يلي:

جدول تحليل التباين (٧ , ١٠) أنوفا (ANOVA)

جدول (٧,١٠)

المصدر	مجموع المربعات SS	درجات الحرية d.f.	متوسط مجموع المربعات MSS	F
بين المعالجات	119.01	2	59.505	$F_0 = 7.84$
الخطأ	166.99	22	7.59	
الكلي	286	24		

من جدول توزيع F آخر الكتاب عند مستوى معنوية $\alpha = 0.01$ نجد أن $F_{2, 22, 0.01} = 5.72$ ويكون فرض العدمي H_0 وهو

$$H_0 = \mu_1 = \mu_2 = \mu_3$$

أو

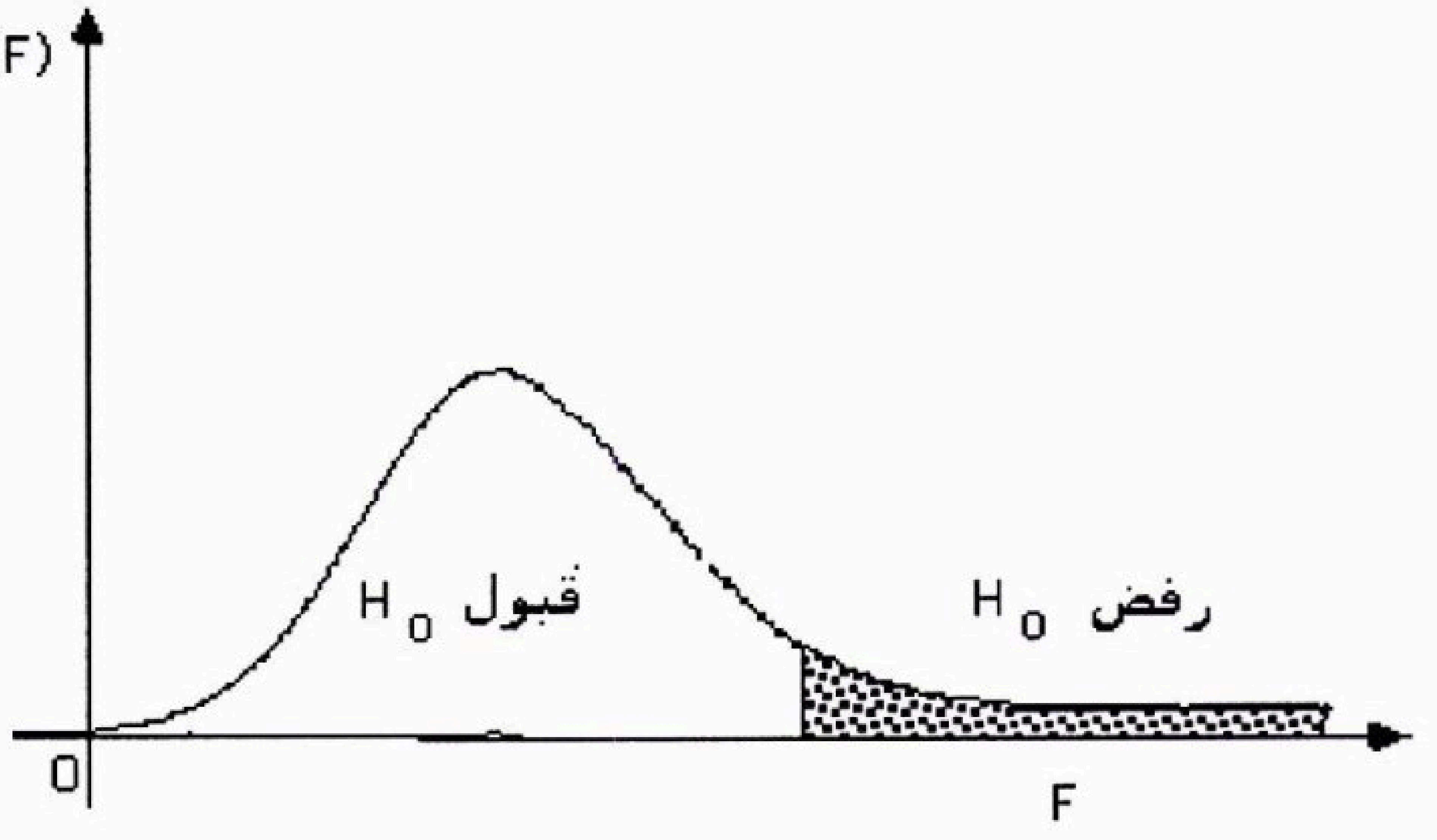
$$H_0 : \tau_1 = \tau_2 = \tau_3 = 0$$

ضد الفرض البديل H_1 هو

على الأقل متوسط واحد مختلف: H_1

ونوضح منطقة قبول H_0، ورفض H_0 بالجزء المظلل بالرسم شكل (٢ , ١٠) كما يلي:

نلاحظ $F_0 = 7.84$ المحسوبة واقعة خارج فترة قبول H_0 وهي $(0, 5.72)$ أي أن نرفض H_0 وهذا يوضح أن الفرق بين طرق التدريس معنوي (مختلف).



شكل (٢ , ١٠)

(٦ , ١٠) نموذج التأثيرات العشوائية

Model of Random Effects

يهتم الباحث ببعض العوامل أو بعامل له عدد كبير من المستويات الممكنة. إذا قام الباحث باختيار بعض هذه المستويات عشوائيا لدراستها فإننا نقول إن هذا العامل عشوائي، ونظرا لسحب المستويات عشوائيا فإن الاستدلال الناتج من تلك التجربة سيكون حول كل مجتمع المستويات المستخدمة لذلك العامل. نفترض أن مجتمع مستويات العامل إما غير محدود أو أنه كبير جدا وبدرجة معقولة ليصبح غير محدود. والنموذج في تلك الحالة هو نفسه النموذج السابق في البند (٣ , ١٠) ولكن فروض النموذج تختلف لأن τ_i , $\in_{ij}$ متغيرات عشوائية ومستقلة، $V(\in_{ij}) = \sigma^2 \ , \ V(\tau_i) = \sigma_\tau^2$. وأيضا فإن اختبار الفروض حول تساوي متوسطات المعالجات يصبح لا معنى له ولكن يصبح الاختبار

$$H_0 : \sigma_\tau^2 = 0$$

$$H_1 : \sigma_\tau^2 > 0$$

ويظل حسابات مجاميع المربعات كما هو في النموذج الثابت وأيضا حساب F_0 ولكن

$$E(MSS_{trt}) = \sigma^2 + n\,\sigma_\tau^2$$

$$E(MSE) = \sigma^2$$

ويلاحظ أن MSE يعتبر تقديرا للمقدار σ^2 وأن $\dfrac{MSS_{trt} - MSE}{n}$ يعتبر تقديرات للمقدار σ_τ^2.

مثال (٣, ١٠)

استخدمت مجموعة من الأفران لتسخين قطع من المعادن في أحد المعامل. كان مفروضا أن جميع الأفران تعمل عند نفس درجة الحرارة. أختيرت ثلاثة أفران عشوائيا وتم تسجيل درجات الحرارة لكل فرن 4 مرات متتابعة وكانت البيانات كما بالجدول رقم (٨, ١٠) كما يلي:

جدول (٨, ١٠)

الفرن	مستويات الحرارة			
	المستوى الأول	المستوى الثاني	المستوى الثالث	المستوى الرابع
1	491.5	498.3	498.1	493.5
2	488.5	484.7	479	477.1
3	490.0	484.8	488.3	473.0

هل هناك فرق معنوي بين درجات الحرارة للأفران المختلفة؟

الحل

لا تختلف خطوات الحل لتكوين جدول تحليل التباين في مثال (٢, ١٠) السابق عن حالة التأثيرات الثابتة. والاختلاف الوحيد هو صياغة الفرض العدمي H_0 ففي هذه الحالة يكون كما يلي:

$$H_0 : \sigma_\tau^2 = 0$$

ضد الفرض البديل

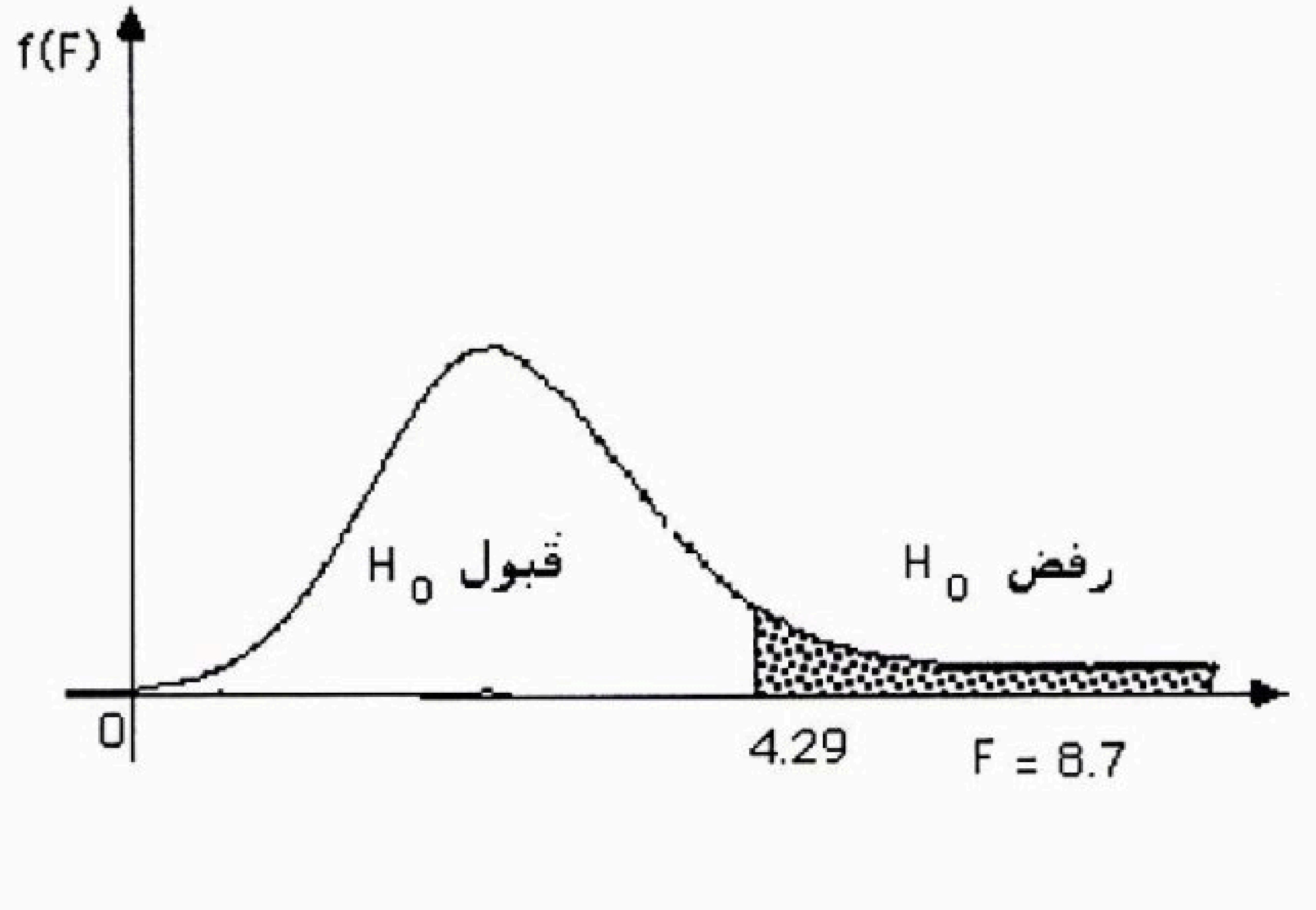
$$H_1 : \sigma_\tau^2 > 0$$

يتم اختبار الفرض العدمي H_0 عند مستوى معنوية α بحساب F_0 من جدول تحليل التباين حيث تلخص النتائج كما هو في جدول (٩, ١٠) كما يلي:

جدول (٩ , ١٠)

المصدر	مجموع المربعات SS	درجات الحرية d.f	متوسط مجموع المربعات MSS	F_0
بين الأفران	439.55	2	219.78	8.7
الخطأ	253.14	9	25.21	
الكلي	674.71	11		

نختار $\alpha = 0.05$ من الجداول F آخر الكتاب نجد أن $F_{2,9,0.05} = 4.29$ وبذلك تكون منطقة رفض H_0 وهي $(0, 4.29)$ كما هو موضح بالجزء المظلل في شكل (٣ , ١٠) كما يلي



شكل (٣ , ١٠)

نلاحظ $F_0 = 8.7$ المحسوبة واقعة خارج منطقة قبول H_0 وبذلك نرفض H_0 وهذا معناه أن الأفران ليست متجانسة تماما في عملية تسخين المعادن.

(٧ , ١٠) تصميم تام التعشية
Completely Randomized Design

يعتبر هذا التصميم مفيدا حينما تكون الوحدات التجريبية متجانسة تجانسا تاما، ويعتبر من أبسط وأسهل التصميمات الإحصائية كما أنه أكثر مرونة عن التصميمات الأخرى. ويستخدم هذا التصميم في حالة تجانس الوحدات التجريبية تماما، وينطبق ذلك على التجارب المعملية، وفي الصوب والحضانات الزراعية، ويقل استخدامه في التجارب الحقلية، كما أنه يستخدم في حالات التجارب الصغيرة والتي لا تحتاج لزيادة الدقة باستخدام التصميمات التي تقلل درجات حرية الخطأ، وأيضا في حالة وجود احتمال لفقدان بعض الوحدات التجريبية. ويعتمد هذا التصميم على توزيع الوحدات التجريبية العشوائية بعشوائية كاملة، أي أن كل وحدة تجريبية لها نفس الفرصة لاستقبال المعالجة وكذا كل معالجة لها فرصة متساوية بالنسبة لكل وحدة تجريبية. ولمعرفة كيف تتم التعشية افترض أن لدينا ثلاث معالجات كل معالجة لها خمس مكررات أي لدينا 15 وحدة تجريبية، في هذه الحالة اعط رقما من 1 إلى 15 لكل وحدة تجريبية، وباستخدام هذه الأرقام العشوائية أختير 15 عددا عشوائيا مكون من ثلاثة أرقام كما هو موضح في جدول (١٠ , ١٠) التالي:

جدول (١٠ , ١٠)

118	701	789	965	683	638	العدد العشوائي
1	8	9	15	7	5	رتبة العدد تصاعديا
901	841	396	802	867	938	العدد العشوائي
13	11	4	10	6	14	رتبة العدد تصاعديا
377	392	848				العدد العشوائي
2	3	12				رتبة العدد تصاعديا

ونرتب الأعداد العشوائية تصاعديا، فنجد أن العـدد 118 أصغر قيمة ويأخـذ الرتبة 1 ونجد أن العدد 965 أكبر رقم ويأخذ الرتبة 15، وهكذا لباقي الأعداد. تأخذ الرتب كما هو موضح اسفل كل رقم عشوائي في الجدول السابق. نجد أن المعالجـة الأولى تأخذ الخمسة أرقام الأولى في الجدول السابـق وهـي 1، 8، 9، 15 و 7 والمعالجة الثانية تأخذ الخمسة التالية وهي 5، 13، 11، 4 و 10 والمعالجة الثالثة تأخذ الخمسة الأخيرة وهي 6، 14، 2، 3 و 12. ينطبق هذا الأسلوب أيضا في حالة ما إذا كانت التكرارات غير متساوية مثلا 6، 6 و فتأخذ المعالجة الأولى ستة أرقام الأولى من الجدول وهي 1، 13، 11، 4 و 10 والمعالجة الثانية ستة أرقام 13، 11، 4، 10، 6 و 10 والمعالجة الثالثة ثلاثة أرقام 2، 3 و 12 كما هو موضح في الجدول (١٠ , ١٠) السابق ونلخص مزايا وعيوب التصميم التام التعشية (Advantage and disadvantage of completely randomized design) فيما يلي:

(١ , ٧ , ١٠) المزايا Advantage

(أ) المرونة حيث يمكن استخدام أي عدد من المكررات أو المعالجات والعامل المحدد فقط هو عدد الوحدات التجريبية المتاحة للتجربة.

(ب) إمكانية استخدام عدد مختلف بين المكررات من معالجة إلى أخرى ولـكـن يستحسن توحيد عدد المكررات.

(ج) سهولة التحليل الإحصائي.

(د) استمرار سهولة التحليل الإحصائي عند فقدان أي مشاهدة.

(هـ) يعتمد هذا التصميم على عدد قليل من الفروض بخلاف التصميمات الأخرى.

(و) عدد درجات الحرية للخطأ التجريبي أكبر من أي تصميم آخر.

(٢ , ٧ , ١٠) العيوب Disadvantage

(ا) عدم كفاءة التصميم في حالة عدم تجانس الوحدات التجريبية.

(ب) دقة التجربة لهذا التصميم منخفضة بالمقارنة بالتصميمات الأخرى لارتـفـاع مقدار الخطأ التجريبي، وعلى الرغم من ارتفاع عدد درجات الحرية للخـطـأ التجريبي إلا أن هذا الارتفاع لا يؤدي إلى تقليل الخطأ التجريبي نفسه، والذي يتسبب فيها القيود المفروضة على العشوائية وخاصة عند وجود عدم التجانس.

مثال (٤ , ١٠)

افترض لدينـا تجربة لدراسة تأثير مستويات سماد معينة على كمية محصول الحبوب، وكان لدينا 6 مستويات من ذلك السماد وكررت كل معالجة 6 مرات من خلال تصميم تام التعشية وكانت المعـالجـات a, b, c, d, e, f حيث المعالجة a بـدون تسميد (مجموعة ضابطة) والمعالجات الخمس الباقية تمثل المستويات المختلفة للسماد حسب جدول (١١ , ١٠) التالي:

جدول (١١ , ١٠)

اسم المعالجة	a	b	c	d	e	f
كمية نترات الكالسيوم للفدان	--	50 kg	100	150	200	250

ووزعت المعالجات عشوائيا على عـدد 36 قطعة بالحقل وكان التوزيع للمعالجـات وكمية الحبوب المنتجة بالكيلوجرام للقطعة حسب جدول (١٢ , ١٠) كما يلي:

جدول (١٢ , ١٠)

6 a	19 e	14 f	7 a	12 b	13 c
10 b	17 d	14 c	18 e	71 e	16 d
15 d	6 a	15 f	17 d	15 f	16 f
14 c	15 c	13 c	7 a	15 f	11 b
11 b	19 e	16 d	6 a	12 b	18 e
15 d	10 b	8 a	17 e	14 c	15 f

والمطلوب إجراء تحليل التباين لمعرفة إذا كان مستويات السماد تأثير معنوي بين متوسطات الإنتاج للحبوب عند مستوى معنوية $\alpha = 0.05$.

الحل

نجري التحليل الإحصائي كما يلي:

لتكن $Y_{i.,\, i=1,2,\dots,6}$ تمثل مجموعة كمية الانتاج للحبوب كل مستوى من المستويات الستة للسماد ومن جدول (١٢ , ١٠) نجد أن

للمستوى a فإن $Y_{1.} = 6+6+8+7+7+6 = 40\ \text{kg}$

للمستوى b فإن $Y_{2.} = 10+11+10+12+12+11 = 66\ \text{kg}$

للسمتوى c فإن $Y_{3.} = 14+15+13+14+14+13 = 83\ \text{kg}$

للسمتوى d فإن $Y_{4.} = 15+15+17+16+17+16 = 96\ \text{kg}$

للمستوى e فإن $Y_{5.} = 19+19+18+17+17+18 = 108\ \text{kg}$

للمستوى f فإن $Y_{6.} = 14+15+15+15+16+15 = 90\ \text{kg}$

وكمية الإنتاج الكلي $Y..$ هي

$$Y.. = \sum_{i=1}^{6} Y_{i.} = 40+66+83+96+108+90 = 483\ \text{kg}$$

ومعامل التصيحح CF هو

$$CF = \frac{(483)^2}{36} = 6480.25$$

ومجموع المربعات الكلي SS_{Tot} هو

$$SS_{Tot} = (6)^2 + (19)^2 + \dots + (14)^2 + (15)^2 - CF$$

$$= 6991 - 6480.25 = 510.75$$

ومجموع المربعات بين المعالجات (المستويات) SS_{trt} هو

$$SS_{trt} = \frac{(40)^2}{6} + \frac{(66)^2}{6} + \frac{(83)^2}{6} + \frac{(96)^2}{6} + \frac{(108)^2}{6} + \frac{(90)^2}{6} - CF$$

$$= 6970.84 - 6480.25 = 490.59$$

ويمكن تلخيص النتائج السابقة في جدول تحليل التباين (١٣ , ١٠) كما يلي:

جدول (١٣ , ١٠)

المصدر / الاختلاف	مجموع المربعات SS	درجات الحرية d.f	متوسط مجموع المربعات MSS	F_0
بين المعالجات (المستويات)	490.59	5	98.12	$F_0 = \frac{98.12}{0.672}$ $= 146.01$
الخطأ	20.16	30	0.672	
الكلي	510.75	35		

ويصاغ الفرض العدمي H_0 حيث إن

$$H_0 : \mu_a = \mu_b = \mu_c = \mu_d = \mu_e = \mu_f$$

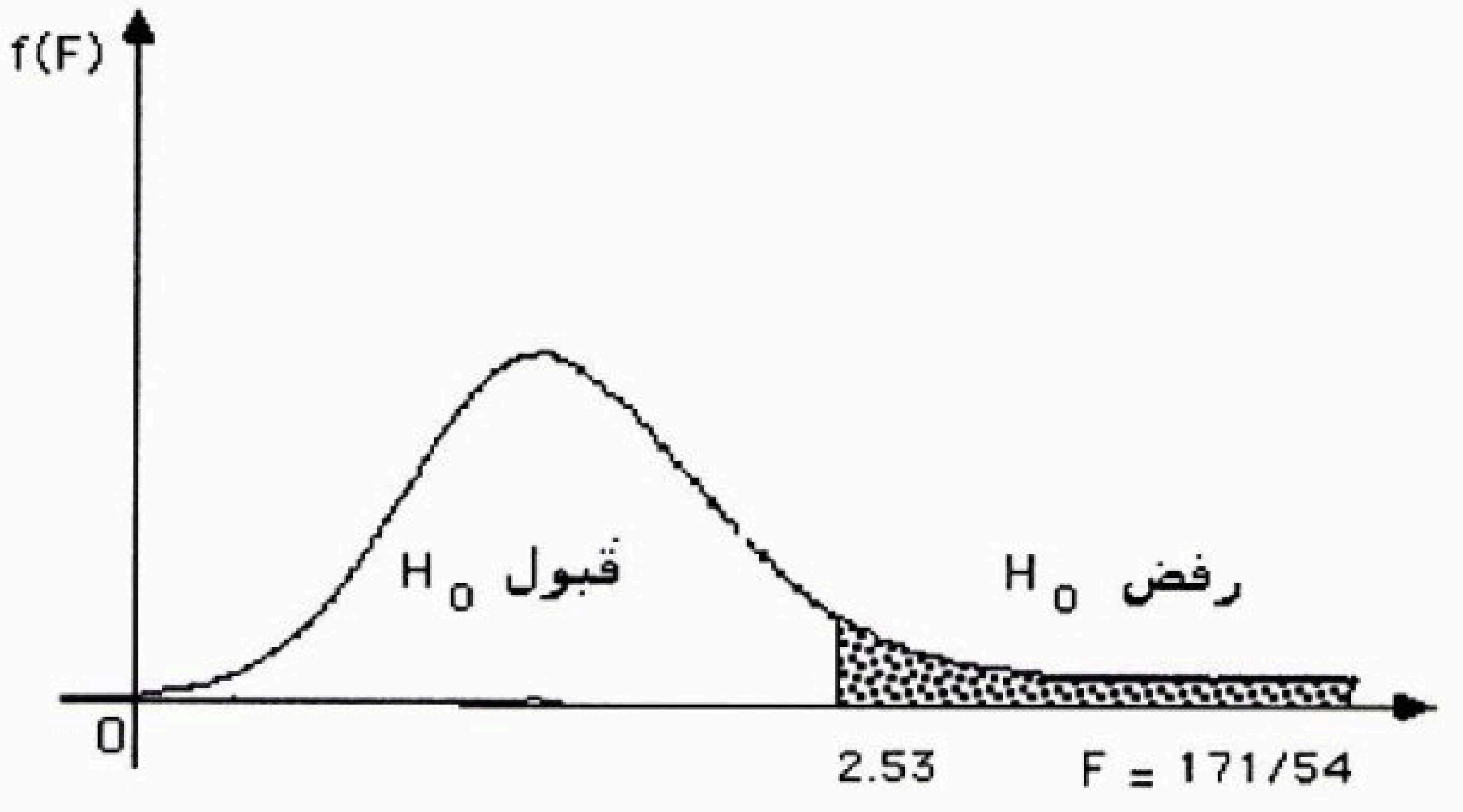
أو

$$\tau_a = \tau_b = \tau_c = \tau_d = \tau_e = \tau_f = 0$$

ضد الفرض البديل H_1 حيث إن

H_1: على الأقل متوسط واحد مختلف

وقيمة $F_{5, 30, 0.05} = 2.53$ من جداول F عند مستوى معنوية $\alpha = 0.05$ ونوضح في شكل (٤ , ١٠) منطقة قبول H_0 (0, 2.53) ومنطقة رفض H_0 بالجزء المظلل بالرسم كما يلي:



شكل (١٠,٤)

وحيث إن $F_0 = 146.01$ واقعة في منطقة رفض H_0 أي أن تأثير مستويات السماد مختلفة (معنوي) بدرجة ثقة 95%.

(١٠,٨) تمارين

١ - طلب من أربعة كميائين تحديد نسبة مثيل الكحول في أحد المركبات الكيميائية، وأعطى لكل كيمائي ثلاثة عينات متشابهة - تم تخصيصها عشوائيا - من المادة الكيميائية لتحليلها، وتم تسجيل النتائج الأربعة بالجدول التالي:

نسبة مثيل الكحول			
العينة الثالثة	العينة الثانية	العينة الأولى	الكيميائي
84.38	84.04	84.99	1
84.88	85.13	85.15	2
84.16	84.48	84.72	3
84.55	84.10	84.20	4

هل الاختلاف بين الكيميائيين الأربعة معنوي؟ ($\alpha = 0.05$) .

٢ - في دراسة لمعرفة تأثير مستويات عقار معين على ضغط الدم، استخدم الباحث 24 فأرا من فئران التجارب متشابهة، وتم تخصيص 6 فئران لكل مستوى من مستويات ضغط الـدم (4 مستويات مختلفة من العقار) وتم قياس الضغـط وكانت النتائج كما يلي:

مستويات العقار بالملجرام			
الرابع 20 ملجم	الثالث 15 ملجم	الثاني 10 ملجم	الأول 5 ملجم
85	86	91	97
88	90	89	91
83	89	88	99
83	91	90	93
86	94	95	91
58	84	93	96

ما نوع التصميم المستخدم في التجربة السابقة؟

حلل البيانـات السـابقة واختبر اختلاف المستويـات السـابقة عند مـسـتـوى معنوية $\alpha = 0.05$.

٣ - أراد أحد مهندسي ضبط الجودة دراسة تأثير الحرارة على عمر نوع معين مـن بطاريات السيارات. تم اختيار 20 بطارية عشوائيا. تم توزيع البطاريات على أربعة مجموعات. تم تعريض كل مجموعة لدرجة حرارة معينة ضمن أربعة مستويات (منخفضة - عادية - عالية - عالية جدا). تم تسجيل عمر فشل البطاريات بالساعات كما هو مبين بالجدول التالي:

مجموعات درجة الحرارة			
عالي جدا	عالي	عادي	منخفض
5.1	6	7.6	8
6.6	6.3	8.2	8.1
5.9	7.1	9.8	9.2
6.7	7.7	10.9	9.4
7.8	8.9	12.3	11.7

(أ) هل تدل هذه البيانات على اختلاف متوسط عمر البطارية عند المستويات المختلفة لدرجة الحرارة عند مستوى معنوية $\alpha = 0.05$؟

(ب) ما نوع التصميم المستخدم؟

٤ - رغب أحد الباحثين في اختبار تأثير الضوضاء على قدرة الشخص على القيادة لمعرفة هل تؤثر زيادة الضوضاء على قدرة الشخص على القيادة. اختير 20 شخصا بنفس القدرة على القيادة، وتم توزيعهم عشوائياعلى أربع مجموعات لمستوى الضوضاءكما يلي:

المستوى الأول للضوضاء 95، المستوى الثانى للضوضاء 100

المستوى الثالث للضوضاء 105، المستوى الرابع للضوضاء 110

وتم عمل اختبار للقيادة وسجلت درجات نتائج الاختبار كما هو موضح بالجدول التالي:

مستوى الضوضاء			
مستوى 95	مستوى 100	مستوى 105	مستوى 110
71	79	69	67
82	75	73	65
78	66	76	63
80	70	78	75
72	77	74	72

أجري تحليلا إحصائيا لمعرفة هل يختلف متوسط القدرة على القيـادة تحـت مستويات الضوضاء المختلفة $(\alpha = 0.05)$.

٥ - تمت دراسة مقاومة الشد لأحد أنواع الأسمنت الناتج من أربعة أسـاليـب للخلط وتم الحصول على النتائج الآتية:

أسلوب الخلط	المقاومة للشد رطل / بوصة			
1	3129	3000	2865	2890
2	3200	3300	2975	3150
3	2800	2900	2985	3050
4	2600	2700	2600	2765

اختبر الفرض القائل أن أساليب الخلط لا تؤثر على مقاومة الشد لذلك النوع من الأسمنت $\alpha = 0.01$.

٦ - شك مدير إنتاج أحد المصانع في أن نسبة الكالسيوم في المادة الخام المرسلة له في صناديق مختلفة من صندوق الى آخر. وجد المدير أن لديه أعدادا كبيرة من الصناديق في المخزن. قام باختيار خمسة صناديق عشوائيا وقام الكيميائي المختص بأخذ خمسة عينات من المادة الخام من كل صندوق. تم تحليل تلك العينات وكانت النتائج كما في الجدول التالي:

الصندوق الخامس	الصندوق الرابع	الصندوق الثالث	الصندوق الثاني	الصندوق الأول
23.29	23.28	23.51	23.59	23.46
23.46	23.40	23.64	23.46	23.48
23.37	23.70	23.46	23.46	23.56
23.32	23.46	23.52	23.49	23.39
23.38	23.39	23.49	23.50	23.40

(ا) هل تعتقد أن شك المدير في محله؟ أي هل هناك اختلاف بين نسبة الكالسيوم من صندوق لآخر $\alpha = 0.05$.

(ب) ما نوع النموذج المستخدم (تأثيرات ثابتة - تأثير عشوائي).

٧ - تمت دراسة تأثير 4 عوامل مساعدة على تركيز أحد مكونات مزيج من سائل مكون من عدة مكونات. وحصل على التركيزات الآتية:

العوامل المساعدة			
العامل الأول	العامل الثاني	العامل الثالث	العامل الرابع
52.9	50.1	56.3	50.2
49.9	54.2	54.5	57.2
50	56.4	57	58.4
51.7	—	55.3	55.8
—	—	—	54.9

هل للعوامل المساعدة الأربعة نفس التأثير على التركيز؟ اعتبر $\alpha = 0.01$.

٨ - لدراسة المقاومة لأربعة أنواع من الأسلاك تم اختيار أربعة أسلاك عشوائيا من كل نوع وقيست المقاومة لوحدة الطول - السلك وسجلت النتائج وتم إجراء الحسابات الآتية: $SSE = 0.22$, $SS_{Tot} = 1.18$

اختبر معنوية الفروق بين متوسطات مقاومة الاسلاك عند مستوى معنوية $\alpha = 0.05$.

٩ -سجلت درجات طلاب ثلاث شعب من شعب 103 إحص وتم عشوائيا اختيار درجات 5 طلاب من الشعبة الأولى ودرجات 7 طلاب من الشعبة الثانية ودرجات 4 طلاب من الشعبة الثالثة. تم حساب جدول تحليل التباين كما هو موضح فيما يلي:

F	متوسط مجموع المربعات	درجات الحرية d.f. الحرية	مجموع المربعات SS	مصدر الاختلاف
4				بين الشعب
	3			الخطأ
				الكلي

(أ) اكمل جدول تحليل التباين

(ب) اختبر معنوية الفروق بين متوسطات درجات الطلاب في شعب الطلاب عند $\alpha = 0.01$.

١٠ - في دراسة لمعرفة الفرق بين تأثير ثلاث طرق مختلفة للتدريس. تم تخصيص مجموعة من الطلاب لكل طريقة عشوائيا وتم اختيار جميع الطلاب بعد فترة وكانت نتائج الاختبارات كما يلي:

الطريقة الأولى	4	5	4	3	6	10	1	
الطريقة الثانية	21	8	10	5	7	9	14	4
الطريقة الثالثة	1	3	4	6	8	5		

حيث كان مجموع مربعات القيم هو$1029 = \sum\sum y_{ij}^{2}$. إجر تحليل التباين المناسب(α=0.05) .

١١-أراد باحث أن يختبر تأثير 4 محاليل كيميائية على قوة نوع معين من القماش فجرب المحاليل على أربعة أتواب وحصل على النتائج التالية:

التوب الرابع	التوب الثالث	التوب الثاني	التوب الأول	المحاليل
1	0.6	0.4	0.3	المحلول الأول
0.9	0.8	0.3	0.4	المحلول الثاني
0.7	0.5	0.4	0.2	المحلول الثالث
1.2	1	0.6	0.7	المحلول الرابع

اجر تحاليل التباين لمعرفة هل الفرق بين تأثير المحاليل معنوي عند ($\alpha = 0.01$)

حيث كان مجموع مربعات القيم هو $\sum\sum x_{ij}^2 = 7.54$.

الفصل الحادي عشر

طرق المقارنات المتعددة

Multiple Comparison Procedures

● مقدمة ● المتضادات ● المتضادات المتعامدة ● اختبار شيفيه ● مقارنة أزواج متوسطات المعالجات ● مقارنة جميع المتوسطات بمتوسط المجموعة الضابطة ● تمارين

(١١,١) مقدمة

Introduction

يستخدم الباحثون في العلوم التجريبية، الاختبارات الإحصائية للحكم على الفرضيات الخاصة بالمتوسطات، وكما رأينا في الفصل السابق اختبار F لاختبار وجود فروق حقيقية بين متوسطات المعالجات. فإذا لم نستطع رفض الفرض العدمي(H_0) فهذا يعني عدم وجود خلاف معنوي بين متوسطات المعالجات، ولا داعي لعمل أي تحليل إحصائي آخر. ومع ذلك يعتبر ذلك في بعض الأحيان تبسيطا يخل بمعنى وقيمة التحليل الإحصائي.

أما إذا رفضنا الفرض العدمي(H_0) باستخدام F. أي أن هناك فروق معنوية بين متوسطات المعالجات تحت احتمال معين، ولكننا لا ندري أي المتوسطات التي تختلف عن المتوسطات الأخرى للمعالجات، ومن الواضح أن أقصى فرق مشاهد بين المتوسطات يكون معنويا. وذلك لأن قيمة F معنوية. لكن ماذا يمكن أن تقول

عن الفروق الأخرى بين المتوسطات؟ لقد اهتم الإحصائيون بدراسة طرق التمييز بين المتوسطات. تسمى طرق التمييز هذه عادة بطرق المقارنات المتعددة. هناك عدة طرق للمقارنات المتعددة سنقوم بدراستها في هذا الفصل فيما يلي.

(٢ , ١١) المتضادات

Contrasts

افترض أنه في تحليل التباين للنماذج الثابتة تم رفض فرض العدم(H_0). كما سبق أن ذكرنا هذا يعني اختلافا معنويا بين المتوسطات، لكنه لا يوضح أي متوسطات تختلف عن بعضها. لذا فإننا نقوم في هذه الحالة بإجراء مقارنة بين المتوسطات المختلفة. كما هو معروف أن $\overline{Y}_{i.}$ هو تقدير للمتوسط μ_i ويمكن استخدام $\overline{Y}_{i.}$ أو $Y_{i.}$ في المقارنات. اعتبر مثال (٤ , ١٠) في الفصل السابق (مستويات السماد) حيث تم رفض فرض العدم($H_0 : \tau_i = 0$)، هذا معناه أن كل أو بعض مستويات السماد تأثر على كمية محصول الحبوب. ولكن أي من تلك المستويات تؤثر تأثيرا معنويا.

على سبيل المثال، قد نشك أنه ليس هناك فرق بين تأثير المستوى 2 (c) وتأثير المستوى 3 (d) أي نريد اختبار الفرض $H_0 : \mu_2 = \mu_3$ ضد الفرض البديل $H_1 : \mu_2 \neq \mu_3$ هذا الفرض يمكن اختباره بدراسة التوفيقة الخطية المناسبة من مجموع المعالجات أي $Y_{2.} - Y_{3.} = 0$. إذا كنا أن نعتقد أن المستوى 2 (c) والمستوى 3 (d) لا يختلفان عن المستوى 4 (e) والمستوى 5 (f). فيكون الفروض بالصيغة التالية: $H_0 : \mu_2 + \mu_3 = \mu_4 + \mu_5$ ضد الفرض البديل $H_1 : \mu_2 + \mu_3 \neq \mu_4 + \mu_5$. ذلك يؤدي إلى التوفيقة الخطية التالية:

$$Y_{2.} + Y_{3.} - Y_{4.} - Y_{5.} = 0$$

وبصفة عامة فإن أي مقارنة تحت الاهتمام تؤدي إلى توفيقة خطـيـة مـن مجموع المعالجات، وهذا ما يعرف بالمتضادة (contrast) أو المقارنة (comparison).

تعريف (١, ١١)

تعرف المقارنة أو المتضادة بالتوفيقة الخطية من مجاميع المعالجات وهي:

$$C = \sum_{i=1}^{a} c_i Y_{i.} \ , \quad \sum_{i=1}^{a} c_i^2 = 0 \qquad (١١, ١)$$

وتستخدم قيمة C في العلاقة (١ , ١١) في التقدير بحدود الثقة كما تستخدم في اختبارات الفروض. للتسهيل يمكن أخذ القيم $c_i \ , \ i = 1,2,..., a$ دائما أعداد صحيحة. ويجب ملاحظـة أن $\sum_{i=1}^{a} c_i = 0$ شرطا أساسيا دائما. ولأي متضـادة درجة حرية واحدة، ولذا يعتبر اختبار t ملائما. يمكن استخدام اختبار F لاختبار المتضادة أيـضـا. حـيـث $\sum_{i=1}^{a} c_i \mu_i = \sum_{i=1}^{a} c_i \tau_i = 0$: H_0 ضد الـفـرض البديل $\sum_{i=1}^{a} c_i \mu_i = \sum_{i=1}^{a} c_i \tau_i \neq 0$: H_1. ويتم حساب مجموع مربعات المقارنة SSc كما يلي:

$$SSc = \frac{\left(\sum_{i=1}^{a} c_i Y_{i.}\right)^2}{n \sum_{i=0}^{a} c_i^2} \qquad (١١, ٢)$$

ملاحظة

في حالة عدم تساوي n لجميع المعالجات فإن العلاقة (١١,١) تصبح على الشكل التالي:

$$C = \sum_{i=1}^{a} c_i Y_{i.} \quad , \quad \sum_{i=1}^{a} n_i c_i^2 = 0 \qquad (١١,٣)$$

كما تصبح مجموع مربعات المقارنات SSc من العلاقة (١١,٣) كما يلي:

$$SSc = \frac{\left(\sum_{i=1}^{a} c_i Y_{i.}\right)^2}{\sum_{i=1}^{a} n_i c_i^2} \qquad (١١,٤)$$

كما أنه يمكن استخدام $\overline{Y}_{i.}$ بدلا من $Y_{i.}$ في العلاقات (١١,١)، (١١,٢)، (١١,٣) و (١١,٤) السابقة.

كما ذكرنا لكل مقارنة درجة حرية واحدة. لإجراء الاختبار لأي متباينة تقارن مجموع مربعاتها (SSc) بمتوسط مجموع مربعات الخطأ، والإحصاءة الناتجة تتوزع F بدرجات حرية 1, N - a حيث $F_{1, N-a} = \frac{SSc}{MSE}$ لاحظ أنه يمكن استخدام توزيع t بدرجات حرية N - a حيث $t_{N-a} = \sqrt{\frac{SSc}{MSE}}$، حيث نلاحظ أن $F_{1, d.f} = t^2_{d.f}$.

(١١,٣) المتضادات المتعامدة

Orthognal Contrasts

تعتبر المتضادات المتعامدة من أهم أساليب المقارنات ويقال عن متضادين للمعاملات c_i , d_i على الترتيب أنها متعامدين إذا كان:

$$\sum_{i=1}^{a} c_i d_i = 0 \qquad (١١,٥)$$

وفي حالة النموذج غير المتزن إذا كان:

$$\sum_{i=1}^{a} n_i c_i d_i = 0 \qquad (١١,٦)$$

لأي مجموعـة من المعالجات a فإنه لدينا a - 1 من المقارنات المتعامدة. يقسم مجموع مربعات المعالجات إلـى a - 1 جزءا مستقلا بين المربعات لكل جزء درجة حرية واحدة. لذا فإن الاختبارات التي تجري على المتباينات مستقلة. يوجد عدة طرق لاختبار معاملات المتضادات المختلفة لأي مجموعة من المعالجات. في الغالب قد تقترح طريقة التجربة أو أهدافها مقارنات معينة.

مثال (١ , ١١)

لو لدينـا a = 3 المعالجة الأولى ضابطة (control) والثانية والثالثة مستويات لعامل معين تحت الدراسة فإن المقارنات المتعامدة الملائمة (يكون لدينا مقارنتين متعامدتـين لكل منهما درجة حرية واحدة) توضح كما في جدول (١ , ١١) كما يلي:

جدول (١ , ١١)

العوامل	المجموعة الضابطة	مستوى 1	مستوى 2
المعاملات	(1)	(2)	(3)
المقارنة الأولى	- 2	1	1
المقارنة الثانية	0	- 1	1

نلاحظ من جدول (١ , ١١) عوامل المقارنـة الأولـى c_i هي 1 ,1 ,2- وكذلك عوامـل المقارنة الثانية d_i هي 1 ,1- ,0 وبالتعويض في العلاقة (٥ , ١١) بالعوامل السابقة نحصل على:

$$\sum_{i=1}^{3} c_i d_i = (-2)(0) + (1)(-1) + (1)(1) = 0 - 1 + 1 = \mathbf{0}$$

إن المقارنة الأولى تقارن تأثير متوسط العامل مع المجموعة الضابطة. أما المقارنة الثانية تقارن المستوى (١) للعامل بالمستوى (٢) للعامل. يجب اختبار معاملات المتباينات قبل إجراء التجربة، ذلك لأن اختبارها بعد إجراء التجربة قد تؤدي بالباحث لاختبار مقارنات يكون الفرق بين المتوسطات كبيرا جدا وقد يكون ذلك الفرق حقيقيا أو يكون راجعا للصدفة أو الخطأ العشوائي.

مثال (٢ , ١١)

في مثالنا (٤ , ١٠) في الفصل السابق (مستويات السماد) كان لدينا ستة مستويات (معالجات) وخمسة درجات حرية. أوجد طرق تجزيء درجات الحرية لخمس درجات حرية منفصلة، أي درجة حرية واحدة لكل مقارنة، والفروض المراد دراستها توضح فيما يلي:

(١) $H_0 : \mu_2 = \mu_3$ أي أن $-Y_{2.} + Y_{3.} = 0$

(٢) $H_0 : \mu_5 = \mu_6$ أي أن $-Y_{5.} + Y_{6.} = 0$

(٣) $H_0 : \mu_2 + \mu_3 = \mu_5 + \mu_6$ أي أن

$$-Y_{2.} - Y_{3.} + Y_{5.} + Y_{6.} = 0$$

(٤) $H_0 : 4\mu_4 = \mu_2 + \mu_3 + \mu_5 + \mu_6$ أي أن

$$Y_{2.} + Y_{3.} - 4Y_{4.} + Y_{5.} + Y_{6.} = 0$$

(٥) $H_0 : 5\mu_1 = \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6$ أي أن

$$-5Y_{1.} + Y_{2.} + Y_{3.} + Y_{4.} + Y_{5.} + Y_{6.} = 0$$

لاحظ أن هذه المقارنات متعامدة.

وإذا اعتبرنا المعالجة (1) بدون سماد وأن المعالجتين (2) و (3) مستوى أول وأن

المعالجة(4) مستوى متوسط وأن المعالجتين (5) و (6) مستوى عالي. فإننا يمكن أن نفهم المقارنات على النحو التالي: الأولى الفرق في داخل المستوى نفسه. والثانية الفرق بين المستوى العالي نفسه. والثالثة الفرق بين المستوى الأول والمستوى العالي. والرابعة الفرق بين المستوى المتوسط والأول والعالي معا. أما الخامسة والأخيرة فهي مقارنة بدون ضد مستويات السماد جميعا. وباستخدام بيانات مثال (٤ , ١٠) السابق وحساب مجاميع المربعات المتعامدة SSc لكل مقارنة يمكننا تلخيص المقارنات في جدول (٢ , ١١) التالي:

جدول (٢ , ١١)

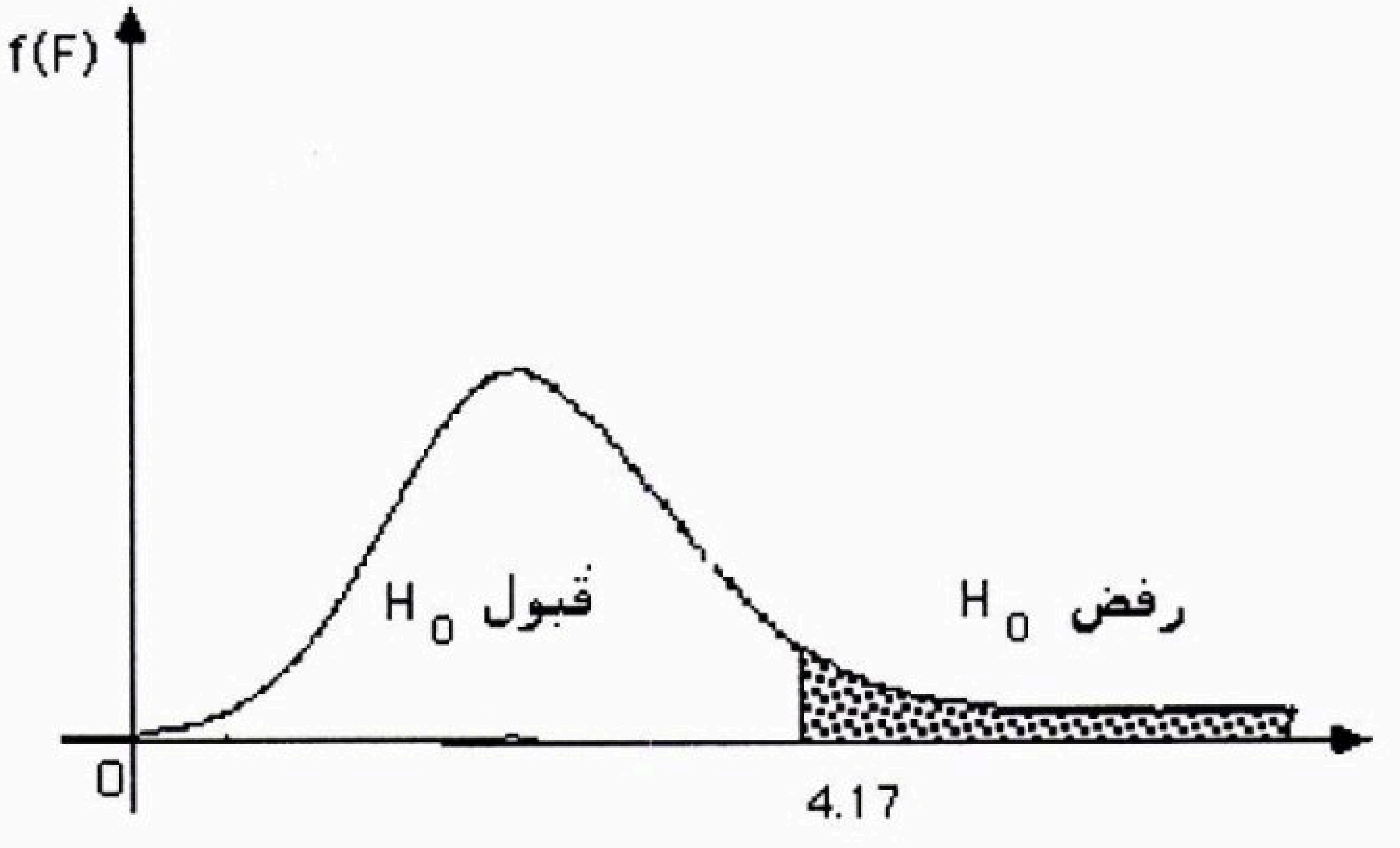
مجموع المعالجات	$Y_{1.}$	$Y_{2.}$	$Y_{3.}$	$Y_{4.}$	$Y_{5.}$	$Y_{6.}$	I	II	SSc
N	40	66	83	96	108	90	$\Sigma c_i Y_{i.}$	$n\Sigma c_i^2$	I^2/II
c_1		-1	1				17	12	24.08
c_2					-1	1	-18	12	27
c_3		-1	-1		1	1	49	24	100.04
c_4		1	1	-4	1	1	-37	120	11.408
c_5	-5	1	1	1	1	1	243	180	328.05

ويمكن صياغة جدول تحليل التباين في جدول (٣ , ١١) كما يلي:

جدول (٣ , ١١)

مصدر الاختلاف	مجموع المربعات SSc_i	درجات الحرية d.f.	متوسط مجموع المربعات MSS	F_0
بين المعالجات	490.59	5	98.12	171.54
c_1	24.08	1	24.08	35.83
c_2	27	1	27	40.18
c_3	100.04	1	100.04	148.87
c_4	11.408	1	11.408	16.98
c_5	328.05	1	328.05	488.17
SSE	20.16	30	0.672	
الكلـــي	510.75	35	--	--

نلاحظ من جدول (١١,٣) أن مجموع مربعات بين المعالجات يساوي مجموع مربعات تلك المقارنات ($\sum_{i=1}^{5} SSc_i = 490.59$)، وأيضا أن جميع المقارنات معنوية حيث $F_{1,30}$ المحسوبة واقعة خارج فترة قبول H_0 عند مستوى معنوية $\alpha = 0.05$ أي أن فترة القبول H_0 هي: ($0, F_{0.05,1,30} = 4.17$) كما هو موضح بالرسم شكل (١١,١) كما يلي:



شكل (١١,١)

وأن معنوية جميع المقارنات غير معروف في جميع الحالات فهناك مقارنات غير معنوية على الرغم من أن F_0 في جدول تحليل التباين (١١,٣) معنوية.

فمثلا في مثال (١٠,١) إذا كانت $C = Y_{1.} - Y_{.3}$ حيث n=9 و MSE=7.35 فإنه باستخدام العلاقة (١١,٢) نجد أن:

$$SSc = \frac{(46 - 34)^2}{2(9)} = 8 \quad , \quad SSc = \frac{SSc}{d.f} = \frac{8}{1} = 8$$

وقيمة F_0 هي $F_0 = \frac{MSSC}{MSE} = \frac{8}{7.35} = 1.09$

وهي قيمة غير معنوية مقارنة بفترة قبول H_0 وهي $(0, F_{1, 24, 0.05} = 4.26)$ أي لا نستطيع رفض H_0 القائل $H_0 : \mu_1 = \mu_3$ على الرغم من معنوية F_0 في جدول تحليل التباين (٣, ١٠).

بصفة عامة في حالة المقارنات المتباينة، فإنه إذا كان لدينا a معالجة يكون لدينا a - 1 درجة حرية لمجموع مربعات المعالجات. وتلك يمكن تجزئتها إلى a - 1 جزء كل جزء له درجة حرية واحدة ومجموع تلك الأجزاء يعطي مجموع مربعات المعالجات المتاحة من البيانات، والواقع أن تقدير ذلك متروك للباحث لتقرير إذا كانت المقارنات لها معنى أم لا، ويتوقف ذلك على طريقة التجربة. إذا لم تكن المقارنات متعامدة فإننا يمكن الحصول على a - 1 مقارنة ولكل منها درجة حرية واحدة ولكن مجموع مربعاتها لا يعطي مجموع مربعات المعالجات.

(٤ , ١١) اختبار شيفيه

Scheffe Test

قد لا يعلم الباحثون في كثير من الحالات أي المقارنات التي يرغبون في مقارنتها أو يهتمون بأكثر من a - 1 مقارنة ممكنة. وقد اقترح شيفيه (1953) طريقة المقارنة أي أو كل المقارنات الممكنة من متوسطات المعالجات، وهذا ما يسمى باختبار شيفيه، وهي طريقة عامة يمكن بها اختبار جميع المقارنات الممكنة، ويمكن استخدامها أيضا لعمل حدود ثقة لأي دالة خطية من المعالم. هذا يعني أن هناك عدد لا نهائي من الاختبارات مسموح إجرائه. وبالضرورة فإن الاختبار سيكون له قيمة حرجة كبيرة لأي مقارنة، وهذا ما يجعله اختبارا مفيدا ومحافظا. ويكون قوة الاختبار منخفضة ويكون الخطأ من النوع الأول على الأكثر α لأي مقارنة ممكنة.

ونحسب القيمة الحرجة لأي مقارنة C كما يلي:

١ - نحسب S_α حيث

$$S_\alpha = \sqrt{(a-1)\, F_{a-1,\, N-a,\, \alpha}} \qquad (١١,٧)$$

٢ - القيمة الحرجة

قيمة شيفيه $S_\alpha\, S_c$ حيث

$$S_c = \sqrt{MSE\left(\sum_{i=1}^{a} c_i^2 / n_i\right)} \qquad (١١,٨)$$

٣ - نرفض فرض العدم

$$H_0 : \sum_{i=1}^{a} c_i\, \mu_i = 0$$

إذا كان $|c| > S_\alpha\, S_c$ وما عدا ذلك نقبل H_0.

مثال (١١,٣)

افترض لدينا لنتائج كما يلي:

$\overline{Y}_{1.}$	$\overline{Y}_{2.}$	$\overline{Y}_{3.}$	$\overline{Y}_{4.}$	$\overline{Y}_{5.}$	$n_i = n = 5$
9.8	15.4	17.6	21.6	10.8	$MSE = 8.06$

افترض أيضا أن

$$H_0 : F_1 = \mu_1 + \mu_3 - \mu_4 - \mu_5 = 0$$

و

$$H_0 : F_2 = \mu_1 - \mu_4 = 0$$

ومن ثم فإن

$$c_1 = \overline{Y}_{1.} + \overline{Y}_{3.} - \overline{Y}_{4.} - \overline{Y}_{5.}$$
$$= 9.8 + 17.6 - 21.6 - 10.8 \quad = 5$$

$$c_2 = \overline{Y}_{1.} - \overline{Y}_{4.}$$
$$= 9.8 - 21.6 = -11.8$$

وباستخدام العلاقة (٨, ١١) نجد أن:

$$Sc_1 = \sqrt{8.06\,(1 + 1 + 1 + 1)/5} = 2.54$$

$$Sc_2 = \sqrt{8.06(1 + 1)/5} = 1.8$$

ولحساب القيمة الحرجة S_α عند $\alpha = 0.01$ هي

$$S_{0.01} = \sqrt{(5-1)F_{0.01,\ 4,\ 20}}$$
$$= \sqrt{4(4.43)} = 4.21$$

ولإجراء الاختبار $H_0 : F_1 = 0$ نرفض H_0 إذ أن

$$|c_1| > S_{0.01}\ Sc_1$$

ولكن $|c_1| = 5$ وأن $S_{0.01}\ Sc_1 = 10.69$

أي أن

$$|c_1| = 5 < S_{0.1}\ Sc_1 = 10.69$$

أي أن لا نستطيع رفض H_0 القائل أن $H_0 : F_1 = 0$ وكذلك إجراء الاختبار $H_0 : F_2 = 0$ نرفض H_0 إذا كان $|c_2| > S_{0.01} \; Sc_2$ ولكن من الحساب نجد أن

$$|c_2| = 1.8 < S_{0.01} \; Sc_2 = 7.58$$

أي أننا لا نستطيع رفض $H_0 : F_2 = 0$.

ملاحظة

يرغب الباحثين في كثير من الحالات العملية مقارنة أزواج من المتوسطات، ويكون الاهتمام بالاختلافات من كل أزواج المتوسطات للمعالجات، ويكون المتضادات على الشكل $F = \mu_i - \mu_j \;\; , i \neq j$. وعلى الرغم من أن طريقة شيفيه يمكن استخدامها في تلك الحالات إلا أنها ليست حساسة لتلك الانواع من المقارنات. وهناك أنواع اخرى من المقارنات يمكن استخدامها في ذلك النوع نذكر منها ما يلي.

(١١, ٥) مقارنةأزواج متوسطات المعالجات

Comparison Between Pairs of Treatment Means

وهذا يعني مقارنة جميع الأزواج الممكنة للمتوسطات المختلفة، أي أننا نهتم بفرض العدم H_0 وهو لكل $H_0 : \mu_i = \mu_j \;, \;\; i \neq j$ حيث $i,j = 1,2,..., a$.

ونستعرض هذه الطرق كما يلي:

(١١, ٥, ١) طريقة أقل فرق معنوي

اقترح فيشر (1935) استخدام أقل فرق معنوي (least significant difference - LSD) للمقارنة بين متوسطين ويحسب كما يلي:

$$LSD = t_{\alpha/2} \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \quad (١١,٩)$$

وفي حالة تساوي $n = n_i = n_j$ فإن

$$LSD = t_{\alpha/2} \sqrt{MSE\left(\frac{2}{n}\right)} \quad (١١,١٠)$$

ويستخدم LSD على افتراض أنه تم رفض فرض العدم H_0 في تحليل التباين وإننا نريد أن نختبر $H_0 : \mu_i = \mu_j$ ضد الفرض البديل $H_1 : \mu_i \neq \mu_j$. في هذه الحالة نقارن قيم LSD بالفرق بين المتوسطات لكل زوج من المتوسطات. إذا كان الفرق أكبر من LSD فإننا نرفض فرض العدم H_0.

مثال (١١,٤)

افترض نفس المثال السابق (١١,٣) لاختبار شيفيه حيث كانت البيانات كما يلي:

$\bar{Y}_{1.}$	$\bar{Y}_{2.}$	$\bar{Y}_{3.}$	$\bar{Y}_{4.}$	$\bar{Y}_{5.}$	$n_i = n = 5$
9.8	15.4	17.6	21.6	10.8	MSE = 8.06

$$LSD(0.05) = t_{0.05}\sqrt{\tfrac{2}{5}(8.06)} = 2.016(1.769) = 3.62$$

$$LSD(0.01) = t_{0.01}\sqrt{\tfrac{2}{5}(8.06)} = 2.845(1.796) = 5.11$$

فالفروق التالية وهي:

$$\bar{Y}_{1.} - \bar{Y}_{5.} = -1$$

$$\bar{Y}_{2.} - \bar{Y}_{3.} = -2.2$$

وهذه الفروق المشاهدة أقل من $LSD(0.05) = 3.62$ فهي غير معنوية أما الفروق التالية وهي:

$$\overline{Y}_{4.} - \overline{Y}_{3.} = 4$$
$$\overline{Y}_{2.} - \overline{Y}_{5.} = 4.6$$

وهذه الفروق المشاهدة أكبـر من $LSD(0.05) = 3.62$ فهـي معنوية عند $\alpha = 0.05$.

أما جميع الفروق المشاهدة السابقة فهـي أقـل مـن $SD(0.01)=5.11$ فهي غير معنوية عند $\alpha = 0.01$ وكطريقة لحساب جميع الفروق الممكنة لتوفير الوقت نـرتـب المتوسطات تصاعديا أو تنازليا ثم نضعها في جدول أفقيا ورأسيا، وكمية الـفـروق الناتجة في جدول (٤, ١١) التالي هو لنفس المثال (٤, ١١) السابق.

ونلاحظ كتابة الجزء العلوي من الجدول (٤, ١١) فقط لأن الجدول متماثـل وهو كما يلي:

جدول (٤, ١١)

	$\overline{Y}_{1.}$	$\overline{Y}_{5.}$	$\overline{Y}_{2.}$	$\overline{Y}_{3.}$	$\overline{Y}_{4.}$
	9.8	10.8	15.4	17.6	21.6
$\overline{Y}_{1.}$	—	−1.0	5.6**	−7.8**	−11.8**
$\overline{Y}_{5.}$		—	4.6*	−6.8**	−10.8**
$\overline{Y}_{2.}$			--	−2.2	−6.2**
$\overline{Y}_{3.}$				—	−4.0*
$\overline{Y}_{4.}$					—

ملاحظة

* الفروق معنوية عند $\alpha = 0.05$ فقط .

** الفروق معنوية عند $\alpha = 0.01$, $\alpha = 0.5$

ويمكن عمل شكل تخطيطي كما يلي :

$$\underline{\overline{Y}_{1.} \quad \overline{Y}_{5.}} \quad \underline{\overline{Y}_{2.} \quad \overline{Y}_{3.}} \quad \overline{Y}_{4.}$$

وهذا معناه أن المعالجة 1 و 5 لا يختلفان معنويا وكذلك المعالجة 2 و 3 وأن المعالجة 4 تختلف معنويا من جميع المعالجات .

ويتردد بعض الإحصائيين في التوصية باستخدام LSD خوفا من سوء استخدامه وخاصة عند عمل مقارنات بعد الحصول على البيانات ، وليست مخططة مسبقا ، أي نحدد المقارنة التي نريد اختبارها قبل التجربة . ويلاحظ أن قيمة α قد تتضخم باستخدام هذه الطريقة وبصيغة محددة حينما α خطأ التجربة من النوع الأول (نسبة عدد التجارب التي يحدث فيها خطأ من النوع الأول إلى العدد الكلي من التجارب) يكون كبيرا ، وهناك بعض الأبحاث المنشورة لحساب قيمة ذلك الخطأ . وقد يفشل LSD في بعض الأحيان في اتخاذ فروق معنوية بين المتوسطات على الرغم من معنوية F والسبب يكمن من أن اختبار F آنيا لجميع أنواع المقارنات ، وليس لمقارنات من الشكل $H_0 : \mu_i - \mu_j = 0$ فقط . وينبغي مراعاة النقط التالية عند استخدام اختبار LSD أي أنه إذا زادت مستويات عامل الدراسة عن 2 فيكون الاختلاف بين المستوى الأكبر والأصغر أكبر من أي متوسطين يختاران عشوائيا . وينبغي عدم استخدام الاختبار حيثما لا يكون F في تحليل التباين معنوية . وأنه يمكن استخدامه في حالة المقارنات المستقلة .

(٢ , ٥ , ١١) اختبار دنكان للمدى المتعدد

أسلوب للمقارنة من أزواج المتوسطات (كل الأزواج الممكنة) تم تقديمه وتطويره بواسطة دنكان ويطلق عليه (Duncans new multiple range) . ويمتاز هذا

الاختبار ببساطة وفيه يجب أن نرتب المعالجات تصاعديا ونقوم بحساب $S_{\bar{Y}_{i.}} = \sqrt{MSE/n_h}$ في حالة عدم تساوي n لجميع المعالجات نحسب $n_h = a/(\sum_{i=1}^{a} n_i)$ وتحسب $S_{\bar{Y}_{i.}} = \sqrt{\frac{MSE}{n_h}}$. ثم نحصل على قيم $r_\alpha(p, f)$ من جدول رقم (٦) آخر الكتاب. حيث α تمثل مستوى المعنوية وأن $\rho = 2,3,\dots, a$ وأن f تمثل درجات حرية الخطأ. يحسب قيمة الإحصاءة $R_\rho = r_\alpha(\rho.f)S_{\bar{Y}_i}$ حيث $\rho = 2,3,\dots, a$ ($R\rho$ تمثل أقصر مدى للمعنوية). يتم اختيار الفروق بين المتوسطات بدءاً من المتوسط الأكبر ضد المتوسط الأصغر ويقارن بالمدى R_a ثم المتوسط الأكبر ضد ثاني أصغر متوسط ونقارن بالمدى R_{a-1} وهكذا. . . حتى يتم المقارنة لجميع الفروق الممكنة والتي عددها $\frac{a(a-1)}{2}$ زوجا من المتوسطات. ونرفض فرض العدم H_0 إذا كان الفرق أكبر من قيمة المدى المتعدد. لاحظ أنه ليس هناك أي فرق معنوي بين أي زوج يقع بين متوسطين لا يختلفان معنويا.

لتوضيح طريقة الاختبار افترض نفس المثال (١١,٣) السابق فإننا نجد أن

$$S_{\bar{Y}_{i.}} = \sqrt{8.06/5} = 1.27$$

ومن الجداول (٦) آخر الكتاب $r_{0.05}(\rho,20),\ \rho = 2,3,4,5$ حيث يتم وضع القيم كما يلي:

ρ	2	3	4	5
$r_{0.05}(\rho, 20)$	2.93	3.10	3.18	3.25

ويمكن حساب

$$R_{\rho} = r_{0.05}\ (\rho,20)\ S_{\overline{Y}_{i.}}$$

ويتم وضع قيم R_{ρ} كما يلي:

ρ	2	3	4	5
R_{ρ}	3.75	3.94	4.04	4.13

ويمكن كتابة فروق المتوسطات $\overline{Y}_{i.}$ ومقارنتها بقيم R_{ρ} كما يلي:

$$\overline{Y}_{4.} - \overline{Y}_{1.} = 21.6 - 9.6 = 11.8 > 4.13 \quad (R_5)$$

$$\overline{Y}_{4.} - \overline{Y}_{5.} = 21.6 - 10.8 = 10.8 > 4.04 \quad (R_4)$$

$$\overline{Y}_{4.} - \overline{Y}_{2.} = 21.6 - 15.4 = 6.2 > 3.94 \quad (R_3)$$

$$\overline{Y}_{4.} - \overline{Y}_{3.} = 21.6 - 17.6 = 4.0 > 3.75 \quad (R_2)$$

$$\overline{Y}_{3.} - \overline{Y}_{1.} = 17.6 - 9.8 = 7.8 > 4.04 \quad (R_4)$$

$$\overline{Y}_{3.} - \overline{Y}_{5.} = 17.6 - 10.8 = 6.8 > 3.94 \quad (R_3)$$

$$\overline{Y}_{3.} - \overline{Y}_{2.} = 17.6 - 15.4 = 2.2 < 3.75 \quad (R_2)$$

$$\overline{Y}_{2.} - \overline{Y}_{5.} = 15.4 - 10.8 = 3.6 < 3.75 \quad (R_2)$$

$$\overline{Y}_{2.} - \overline{Y}_{1.} = 15.4 - 7.8 = 4.6 > 3.94 \quad (R_3)$$

نلاحظ أن هناك فرق معنوي من جميع الأزواج ما عدا الأزواج (3,2) و (2,5) ويمكن تلخيص نتائج المثال السابق كما يلي:

$\underline{9.8 \quad 10.8} \quad \underline{15.4 \quad 17.6} \quad 12.6$

ويتوافر في اختبار دنكن هذا سهولة الحساب كما أن له ميزة أخرى وهو

ارتفاع احتمال رفض معنوية الخلافات بين المتوسطات حينما لا يكون هذه الخلافات حقيقية. ويعتبر ذلك حماية من ارتكاب خطأ من هذا النوع. لذا فإن اختبار دنكن اكثر قوة وتأثيرا في اكتساب الفروق بين المتوسطات عندما تكون تلك الفروق حقيقية. لذا فهو شائع الاستعمال.

(٣, ٥, ١١) اختبار ستودنت نيومان كيلز أو اختبار S.N.K

سمي هذا الاختبار باسم ستودنت - نيومان - كيلز **Student - Newman Keuls** نسبة لثلاثة أشخاص ويطلق عليه أيضا نيومان كيلز أو ببساطة طريقة كيلز، نفس طريقة الحساب هي نفسها طريقة الحساب في حالة دنكان حيث نحسب

$$K_\rho = q_\alpha(\rho,f)\, S_{\overline{X}_{i.}} \qquad (١١,١١)$$

حيث $q_\alpha(\rho,f)$ هي قيمة الحد الأعلى للمدى المعياري عند α لمجموعة متوسطات من حجم ρ و f ودرجات حرية الخطأ وتحسب تلك القيم $q_\alpha(\rho,f)$ من جدول رقم (٦) آخر الكتاب ويعرف بالمدى المعياري

$$q = \frac{\overline{Y}_{max} - \overline{Y}_{min}}{\sqrt{MSE / n}} \qquad (١١,١٢)$$

حيث

$\overline{Y}_{max}$ قيمة أكبر متوسط

$\overline{Y}_{min}$ قيمة أصغر متوسط في تلك المجموعة المكونة من ρ متوسطا.

نحسب K_ρ وبنفس أسلوب دنكان ونستمر. يعتبر هذا الاختبار أكثر تحفظا من اختبار دنكان ومعدل الخطأ من النوع الأول أصغر كما أن قوة الاختبار أقل. لاحظ أن قيمة $q_\alpha(\rho,f) > r_\alpha(\rho,f)$ وهذا معناه أنه من الصعب الحصول على

معنوية زوج من أزواج المتوسطات باستخدام هذا الاختبار عما هو عليه في حالـة اختبار دنكان.

وباستخدام نفس بيانات المثال (٣,١١) السابق يمكن حســاب $q_{0.05}(\rho,20)$ من العلاقة (١١,١١) ونوضحها كما يلي:

ρ	2	3	4	5
$r_{0.05}(\rho, 20)$	2.95	3.10	3.18	3.25
$q_{0.05}(\rho, 20)$	2.95	3.58	3.96	4.24

ويمكن حساب K_ρ لنفس بيانات مثال (٣,١١) من العلاقة (١١,١١) كما يلي:

ρ	2	3	4	5
K_ρ	3.75	4.55	5.03	5.38

نحسب الفرق بين المتوسطات لأكبر ولأصغر قيمة K_ρ كما يلي:

$\overline{Y}_{4.} - \overline{Y}_{1.} = 21.6 - 9.6 = 11.8 > 5.38$ K_5

$\overline{Y}_{4.} - \overline{Y}_{5.} = 21.6 - 10.8 = 10.8 > 5.03$ K_4

$\overline{Y}_{4.} - \overline{Y}_{2.} = 21.6 - 15.4 = 6.2 > 4.55$ K_3

$\overline{Y}_{4.} - \overline{Y}_{1.} = 21.6 - 17.6 = 4.0 > 3.75$ K_2

$\overline{Y}_{3.} - \overline{Y}_{1.} = 17.6 - 9.8 = 7.8 > 5.03$ K_4

$\overline{Y}_{3.} - \overline{Y}_{5.} = 176 - 10.8 = 6.8 > 4.55$ K_3

$\overline{Y}_{3.} - \overline{Y}_{2.} = 17.6 - 15.4 = 2.2 < 3.75$ K_2

$\overline{Y}_{2.} - \overline{Y}_{5.} = 15.4 - 10.8 = 4.6 > 4.55$ K_3

$\overline{Y}_{5.} - \overline{Y}_{1.} = 10.8 - 9.6 = 1.2 < 3.75$ K_2

ويمكن تلخيص النتائج كما في الشكل:

$\underline{9.8 \quad 10.8} \quad \underline{15.4 \quad 17.6} \quad 21.6$

نفس النتائج كما في اختبار دنكان (٢ , ٥ , ١١).

(٤ , ٥ , ١١) اختبار توكي

يعتمد اختبار توكــي **Tukey's procedure** على استخدام المدى المعياري، ويطــبــق للمقارنات بين أزواج متوسطات المعالجات، ويعطى قيمة واحدة للمقارنة لجميــع أزواج المقارنة. لذا فهو سريع وسهل الاستخدام، حيث إن قيمة الاختبار واحــدة لجميع المقارنات فإن تلك القيمة دائما أصغر من القيمة المحسوبة في اختبار شيفيه ويحسب تلك القيمة الحرجة

$$T_{\alpha} = q_{\alpha}(a, n - a)\, S_{\bar{X}_{i.}} \qquad (١١ , ٣)$$

وتستخدم هذه القيمة لجميع المقارنات فمثلا

$$q_{0.05}(5, 20) = 4.23$$

من جدول (٦) آخر الكتاب

$$T_{0.05} = (4.23)\,(1.27) = 5.37$$

وباستخدام بيانات مثال (٣ , ١١) نجد أن $\bar{Y}_{3.}$ و $\bar{Y}_{4.}$ لا يختلفان معنويا وكذلك $\bar{Y}_{2.}$ و $\bar{Y}_{5.}$ بالإضافة إلى الفروق الباقية. ويمكن رسم الشكل التالي:

$$\bar{Y}_{1.} \quad \bar{Y}_{5.} \quad \underline{\bar{Y}_{2.} \quad \bar{Y}_{5.}} \quad \bar{Y}_{4.}$$

نلاحظ صعوبة تفسير تلك النتائج عن سابقتها وأن الاختبار أكثر تحفظا من الاختبارات

السابقة وأن α أقل وقوة الاختبار أيضا أقل.

(١١,٥,٥) أي الاختبارات تستخدم؟

من الصعب الإجابة على هذا السؤال فليست المقارنات السابقة هي كل أنواع الاختبارات. فهناك البعض لم نذكره ولكن في بعض الدراسات وجد أن اختبارLSD أكثر قوة وذلك عند ثبوت معنويــة F في جدول تحليل التباين من غيره. وكذلــك اختبار دنكان وكلاهما موجود في أي حزمة من حزم البرامج الجاهزة الإحصائيــة للكمبيوتر.

(١١,٦) مقارنة جميع المتوسطات بمتوسط المجموعة الضابطة
Comparison All Means with Mean of Control Group

يكون الهدف من التجربة في بعض الأحيان معرفة المعالجات التــي يــكــون تأثيرها أفضل من تأثير معالجة مناسبة، وليس الغرض مقارنة المعالجات ببعضهــا، وتبرز ذلك الفرض لتجارب أخرى. وإذا افترضنا أن المعالجة القياسية أو الضابطة هي رقم 1. فيكون الاختبار

$$H_0 : \mu_1 = \mu_i,\ \ i = 1, 2, \cdots, \alpha$$

ضد الفرض البديل

$$H_1 : \mu_1 \neq \mu_i$$

وهذه الفروض والمقارنات الناتجة عنها لا تكون مستقلة عن بعضها. قــام دانــيــت Duunett's بتعديل اختبار t العادي وحصل على مايسمى باختبار دانيت وله جداول (٧) آخر الكتاب. وهو أسلوب يعطي قيمة واحدة للحكم على المعنوية للفرق بين متوسط كل معالجة ومتوسط المجموعة الضابطة والجدول يمكننا من إجراء اختبار ذو ذيل واحد أو ذيلين. والقيم الجدولية $(d_\alpha (a\text{-}1), f)$ حيث α مستوى العنوية المطلوب،

f درجات حرية الخطأ وقيمة الإحصاءة d' وهي:

$$d' = d_{\alpha}(a-1, f)\sqrt{MSE\left(\frac{1}{n_a} + \frac{1}{n_i}\right)} \qquad (11,14)$$

على فرض أن المعـالجـة a هي المعالجة الضابطة نقارن هذه الـقـيـمـة بـالـفـرق بين$|\overline{Y}_i - \overline{Y}_a|$ فإذا كان $|\overline{Y}_{a.} - \overline{Y}_{i.}| > d'$ نرفض فرض العدم H_0 حيث إن

$$H_0 : \mu_i = \mu_a$$

مثال (١١,٥)

افترض مثال (١١,٣) السابق. اعتبر المعالجـة رقـم 5 هي الضابطة وبذلك يكون لدينا

$$f = 20,\ a - 1 = 4\ ,\ a = 5$$

من الجداول (٧) آخر الكتاب قيمة $d_{0.05}(4, 20) = 2.65$

وتصبح قيمة d' كما يلي:

$$d' = 2.65\sqrt{\frac{2(8.06)}{5}} = 4.76$$

$$\overline{Y}_1 - \overline{Y}_5 = -1 < 4.76$$

$$\overline{Y}_2 - \overline{Y}_5 = 4.6 < 4.76$$

$$\overline{Y}_3 - \overline{Y}_5 = 6.8 > 4.76$$

$$\overline{Y}_4 - \overline{Y}_5 = 10.8 > 5.76$$

الفـرق بـين$(\overline{Y}_3\ ,\ \overline{Y}_5)$ و $(\overline{Y}_4\ ,\ \overline{Y}_5)$ فرق معنوي أي أننـا نـسـتـنـتـج أن $\mu_5 \neq \mu_3$ وكذلك $\mu_5 \neq \mu_4$ ويجب عند إجراء تلك المقارنة استخدام n_a كبيرة أي تكون n كبيرة للمجموعة الضابطة.

(١١,٧) تمارين

١ - إذا كان لديك البيانات الآتية :

ρ رقم المعالجة	1	2	3	4	5	6
$\overline{Y}_{i.}$ المتوسط	4.05	7.76	7.85	10.48	10.54	11.28

لكل متوسط تم حسابه من 12 مشاهدة وكان MSE = 10.697 ودرجات حرية 66 درجة حرية . احسب LSD لجميع أزواج المتوسطات وأيضا استخدام اختبار دنكان .

٢ - بين أن المقارنات التي استخدمت في مثال (١١,٢) السابق متعامدة لكل زوج من المقارنات .

٣ - استخدام بيانات التمرين (١) وقم بتكوين تلك المقارنات

(i) 1 VS 5
(ii) 2,3 VS 4,5
(iii) 2,5 VS 3,4
(v) 2,5 VS 3,4,6
(iv) 4,5 VS 2,3,6

اعمل جدول موضحا قيمة مجموعات مربعات كل معالجـة و F هل تشكل هذه المقارنات مقارنات متعامدة .

٤ - في مثال (١١,٢) استخدم اختبار شيفيه لاختبار الفروض الموجودة في مثال (١١,٢) السابق .

الفصل الثاني عشر

تحليل التباين في اتجاهين
Two Way Analysis of Variance

● مقدمة ● تصميم القطاعات العشوائية الكاملة ● النموذج الرياضي والتحليل الإحصائي ● المقارنات المتعددة ● تمارين

(١٢,١) مقدمة
Introduction

يعتبر استخدام التصميم التام التعشية الذي سبق الكلام عنه في الفصل الحادي عشر السابق مناسبا حينما لا يكون هناك أي مصدرا للاختلاف معروف سوى الاختلاف الناتج عن تأثير المعالجات. قد يكون من المعروف أيضا مسبقا في كثير من الحالات أن بعض الوحدات التجريبية تعطي نتائج مختلفة تحت نفس المعالجة. من أمثلة ذلك في التجارب الحقلية القطع المتجاورة من الأرض لها نفس الاستجابة للمعالجات عن تلك التي تبعد عن بعضها البعض. وبالمثل في الحيوانات التي لها نفس العمر تستجيب الحيوانات التي تثقل وزنا بطريقة مختلفة عن تلك الأخف وزنا. كما أن المشاهدات المأخوذة في يوم قد تختلف عن المأخوذة في يوم آخر. واستخدام آلة معينة قد يؤثر على الوحدات التجريبية بطريقة مختلفة عن آلة أخرى من نفس النوع.

في مثل تلك الحالات التي تم استعراضها وبالمثل في حالات أخرى كثيرة مشابهة يمكن تقسيم الوحدات التجريبية طبقا للمصادر التي قد يكون فيها اختلاف نسبيا. فمثلا في التجارب الحقلية مثلا تقسم الأرض إلى أقسام كل قسم يحتوي

على مجموعة من القطع المتجاورة وكذلك في تجارب الحيوانات يمكن وضع الحيوانات ذات الوزن الثقيل في مجموعة وذات الوزن الخفيف في مجموعة أخرى. . . الخ. وأيضا يمكن وضع إنتاج اليوم الأول في مجموعة وإنتاج اليوم الثاني في مجموعة وهكذا. وبالنسبة للآلات يمكن وضع إنتاج الآلة الأولى في مجموعة والآلة الثانية في مجموعة أخرى. . . الخ. مع مثل تلك الحالات السابقة يمكن تكوين تصميمات تمكننا من فصل الاختلاف الناتج عن تلك المصادر، ويستبعد ذلك الاختلاف عن الخطأ التجريبي. يصبح لدينا في هذه الحالة تحليل تباين متعدد الاتجاهات وذلك لأنه لدينا أكثر من مصدر للاختلافات، ومن ثم تجزئة مجموع المربعات الكلي إلى عدة أجزاء وليس جزئين فقط كما سبق في حالة تحليل التباين في اتجاه واحد.

سندرس في هذا الفصل الحالة حينما يكون هناك مصدرين للاختلاف فقط ويسمى تحليل التباين في اتجاهين. من أهم التصميمات ذات المصدرين للاختلاف هو تصميم القطاعات العشوائية الكاملة. سنقوم بدراسة ذلك التصميم والنموذج الرياضي في حالة التأثيرات الثابتة وبدون تفاعل والتحليل الإحصائي له (تحليل التباين في اتجاهين) كما يلي.

(٢ , ١٢) تصميم القطاعات العشوائية الكاملة

The Randomized Complete Block Design (RCBD)

يعتبر تصميم القطاعات العشوائية الكاملة (RCBD) من أوسع التصميمات استخداما في كثير من التجارب خاصة البسيطة منها. يتم استخدامه حينما يكون لتقسيم الوحدات التجريبية معنى وكل مجموعة تسمى قطاعا. أي يعتمد هذا التصميم على أن الوحدات التجريبية مقسمة في مجموعات متجانسة (قطاعات) ويتم توزيع المعالجات داخل القطاعات بطريقة عشوائية، والغرض من ذلك التجميع هو وضع الوحدات التجريبية المتشابهة تقريبا (المتجانسة) في نفس القطاع ليصبح الاختلاف بين المعالجة والأخرى في داخل القطاع راجع إلى المعالجة فقط. ويصبح

الاختلاف بين الوحدات بين القطاعات المختلفة أكبر في المتوسط من الاختلاف الناتج بين الوحدات داخل القطاع. ويجب حتى ينجح التصميم أن يكون الاختلاف بين متوسطات القطاعات معنويا. وهذا يعني أن التقسيم إلى قطاعات نجح في تقليل جزء من الخطأ التجريبي وأنه لازم للتجزئة.

التقسيم إلى قطاعات شائع وسهل ويمكن ملاحظة المتغيرات التي يمكن التقسيم لها بسهولة. فمثلا في التجارب الحقلية يتكون كل قطاع من قطع مربعة متجاورة من الأرض، وبالمثل في تجارب الحيوانات يتم تقسيم الحيوانات طبقا لأوزانها أو أنواعها أو طرق ترتيبها أو عمرها.. الخ. هناك متغيرات أخرى كثيرة مثل الزمن، الصفات الجغرافية، الآلة المستخدمة، العمر، النوع، الخبرات، التعليم، حجم الوحدات التجريبية... الخ. كل الأمثلة السابقة تصلح كمتغيرات يمكن التقسيم إلى قطاعات طبقا لها. ويتم التعشية (العشوائية) في هذا التصميم داخل كل قطاع.

افترض أن لدينا خمسة معالجات ومتاح لنا أربعة قطاعات، معنى هذا يكون لدينا عدد $5 \times 4 = 20$ وحدة تجريبية. فإننا نوزع المعالجات عشوائيا داخل كل قطاع باستخدام أي طريقة كما تم فيما سبق في تصميم التام التعشية في الفصل العاشر. فمثلا لو فرضنا أن المعالجات كانت هي a, b, c, d, e فيكون مثلا لدينا التوزيع كما في جدول (١, ١٢) كما يلي:

جدول (١, ١٢)

القطاع الرابع	القطاع الثالث	القطاع الثاني	القطاع الأول
e	d	b	a
c	c	a	c
d	a	c	b
a	b	e	e
b	e	d	d

نلاحظ في جدول (١ , ١٢) أن كل معالجة تظهر في كل قطاع مرة واحدة، وأن جميع المعالجات تظهر في كل قطاع. في الواقع كلمة كاملـة (complete) في التسمية للتصميم تشير إلى أن كل قطاع يحتوي على جميع المعالجات.

ونذكر فيما يلي بعض مميزات تصميم القطاعات العشوائية الكاملة.

أولا: سهولة تنفيذ التجربة وسهولة التحليل الإحصائي.

ثانيا: تقليل كمية الخطأ التجريبي باستخدامه عن الحالة التي يستخدم فيها تصميم التام التعشية.

ثالثا: يمكن استخدام أي عدد من المكررات وأي عدد من المعالجات.

من عيوب تصميم القطاعات العشوائية الكاملة نذكر ما يلي:

أولا: حينما تزداد المعالجات يرتفع الاختلاف داخل كل قطاع مما يؤدي إلى زيـادة مقدار الخطأ التجريبي، وفي هذه الحالة ينصح باستخدام التصميمات الأخرى.

ثانيا: درجات الحرية الخطأ أقل بالمقارنة على ما هي عليه في تصميـم الـتـام التعشية.

(٣ , ١٢) النموذج الرياضي والتحليل الإحصائي

Mathematical Model and Statistical Analyses

نفترض أنه لدينا عدد a معالجة يراد مقارنة تأثيرها وذلك في عدد b قطاع. يكون شكل التصميم بصورة عامة كما هو موضح في جدول (٢ , ١٢) التـالـي، حيث يوجد مشاهدة واحدة لكل معالجة في كل قطاع.

كما سبق أن ذكرنا أن توزيع المعالجات على القطاعات يتم عشوائيـا وهـذا مصدر العشوائية الوحيد. والنموذج الرياضي لذلك التصميم يعطى بالعلاقة (١ , ١٢) التالية:

جدول (٢ , ١٢)

القطاعات المعالجات	قطاع 1	قطاع 2	...	قطاع b	المجموع
.	y_{11}	y_{12}	...	y_{1b}	$Y_{1.}$
.	y_{21}	y_{22}	...	y_{2b}	$Y_{2.}$
.	:	:		:	:
.	:	:		:	:
	.	y_{a1}	y_{a2}	... y_{ab}	$Y_{a.}$
المجموع	$Y_{.1}$	$Y_{.2}$		... $Y_{.b}$	$Y_{..}$

$$Y_{ij} = \mu + \tau_i + B_j + \in_{ij} \quad \begin{cases} i = 1,2,\dots, a \\ j = 1,2,\dots, b \end{cases} \qquad (١٢ , ١)$$

حيث μ المتوسط العام، τ_i تأثير المعالجة رقم i و B_j تأثير القطاع رقم j، $\in_{ij}$ الخطأ العشوائي ويتوزع توزيع طبيعي بمتوسط صفرا وتبايـن σ^2 أي $N(0, \sigma^2)$. وأن المعالجات والقطاعات لها تأثيرات ثابته وليـس هـنـاك أي تـفـاعـل فـيـهـا وأيضـا $\sum_{i=1}^{a} \tau_i = 0$, $\sum_{j=1}^{b} B_j = 0$ وفي هذا النموذج نهتم باختـبـار تـسـاوي متوسطات المعالجات أي أن الفرض العدمي H_0 وهو

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_a$$

ضد الفرض البديل

$$H_1 : \mu_i \neq \mu_j \text{ على الأقل واحدة } i, j$$

وهذا يكافيء

$$H_0 : \tau_1 = \tau_2 = ... = \tau_a = 0$$

$H_1 : \tau_i \neq 0$ على الأقل واحده i

والرمز y_{ij} يمثل قيمة المشاهدة الخاصة بالمعالجة رقم i في القطاع رقم j. والرمز $Y_{i.}$ يمثل مجموع المشاهدات الخاصة بالمعالجة رقم i. والرمز $Y_{.j}$ يمثل مجموع المشاهدات الخاصة بالقطاع رقم j. والرمز $Y_{..}$ يمثل مجموع المشاهدات كلها في التجربة محل الدراسة والتي عدد المشاهدات فيها هو $a\,b = N$ مشاهدة. ومن الرموز السابقة يمكن حساب القيم التالية.

متوسط المعالجة i هو

$$\overline{Y}_{i.} = \sum_{j=1}^{b} y_{ij} / a$$

متوسط القطاع j هو

$$\overline{Y}_{.j} = \sum_{i=1}^{a} y_{ij} / b$$

المتوسط الكلي

$$\overline{Y}_{..} = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij} / N$$

يحسب مجموع المربعات الكلي SS_{Tot} كما يلي:

$$SS_{Tot} = \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \overline{Y}_{..})^2$$

أي أن

$$SS_{Tot} = b\sum_{i=1}^{a}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 + a\sum_{j=1}^{b}\left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij} - \overline{Y}_{.j} - \overline{Y}_{.i} + \overline{Y}_{..}\right)^2 \quad (١٢,٢)$$

وتمثل العلاقة (١٢,٢) تجزيئا لمجموع المربعات الكلي SS_{Tot} حيث يمكن إعادة كتابتها بالعلاقة (١٢,٣) كما يلي:

$$SS_{Tot} = SS_{trt} + SS_{blks} + SSE \quad (١٢,٣)$$

وتوضح درجات الحرية لكل مجموع مربعات كما يلي:

(ا) حيث لدينا N مشاهدة فإن درجات الحرية لمجموع المربعــات الكلي SS_{Tot} هي N - 1 درجة حرية.

(ب) حيث إن لدينا a معـالجة فإن درجات الحرية لمجموع مربعات بين المعالجات SS_{trt} هي a - 1 درجة حرية.

(ج) حيث إن لدينا b قطاع فإن درجـات الحرية لمجموع مربعات بين القطاعات SS_{blks} هي b - 1 درجة حرية.

(د) وتصبح درجات الحرية لمجموع مربعات الخطأ SSE هي (a-1)(b-1).

(هـ) يحسب متوسط مجموع مربعات بين المعالجات SS_{trt} بقسمة مجموع مربعات بين المعالجات على درجات حريته (a-1) أي أن

$$MSS_{trt} = SS_{trt}/(a-1) \quad (١٢,٤)$$

(و) كذلك يحسب متوسط مجموع مربعات بين القطاعات SS_{blks} بقسمة مجموع مربعات بين القطاعات على درجات حريته (b-1) أي أن

$$MSS_{blks} = \frac{SS_{blks}}{b-1} \quad (١٢,٥)$$

(ز) وأخيرا يحسـب متوسط مربعات الخطأ MSE بقسمة مجموع مربعات الخطأ على درجات حريته (a-1)(b-1) فيصبح

$$MSE = SSE / (a-1)(b-1) \quad (١٢,٦)$$

لاحظ أن توقعات كل من MSE ، MSS_{blks} و MSS_{trt} كما يلي:

$$E(MSS_{trt}) = \sigma^2 + b\sum_{j=1}^{b} \tau_i^2 / (a-1)$$

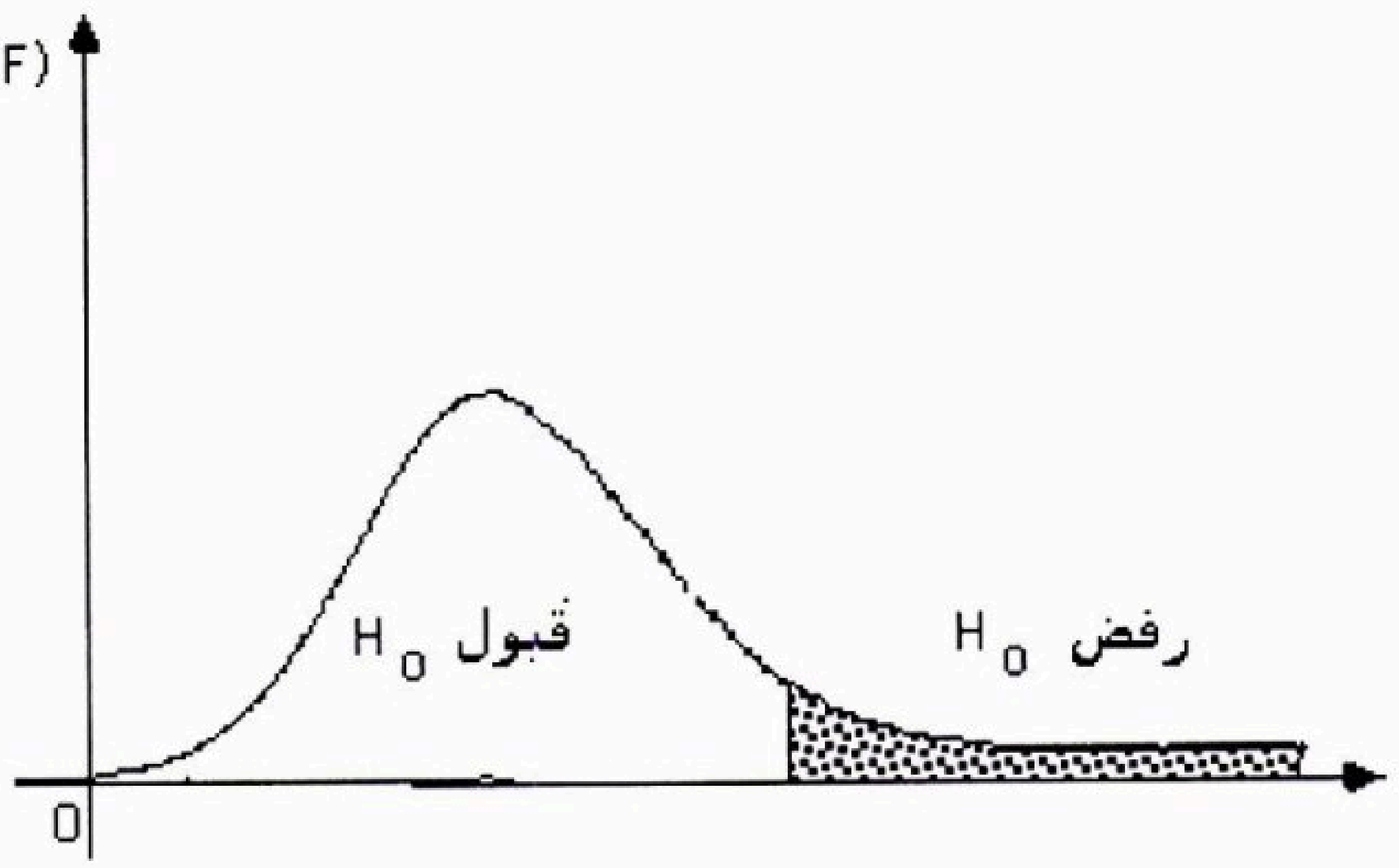
$$E(MSS_{blks}) = \sigma^2 + a\sum_{j=1}^{b} B_j^2 / (b-1)$$

$$E(MSE) = \sigma^2$$

ولذا لاختبار صحة الفرض العدمي H_0، أي تساوي المتوسطات تستخدم الإحصاءة $F_{(a-1),(a-1)(b-1)}$ ويرمز لها بالرمز F_0 وهي

$$F_0 = \frac{MSS_{trt}}{MSE}$$

حيث فترة قبول H_0 عند مستوى معنوية α هي $\left[0, F_{(a-1),(a-1)(b-1),\alpha}\right]$ وتوضح منطقة قبول H_0 مع فترة رفض H_0 بالجزء المظلل في شكل (١٢,١) كما يلي:



شكل (١٢,١)

ولمعرفة ما إذا كان التقسيم إلى قطاعات ذو فائدة أي أن الفرق بين متوسطات القطاعات معنوي. فإننا نختبر الفرض العدمي H_0 ضد الفرض البديل H_1 عند مستوى معنوية α حيث

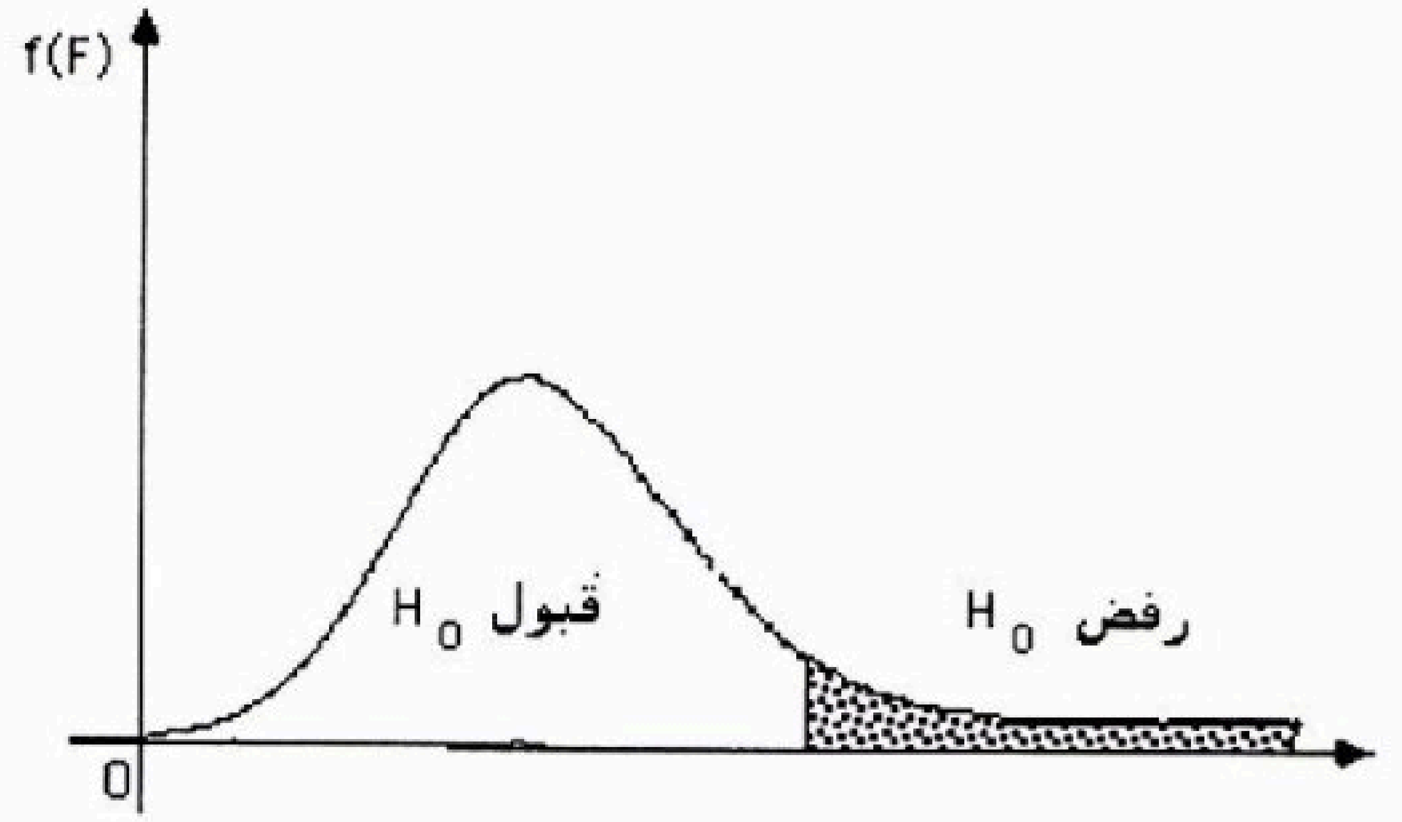
$$H_0 : B_j = 0 \quad ; \quad j = 1,2,\ldots, b$$

$H_1 : B_j \neq 0$ واحدة على الأقل j

نحسب الإحصاءة $F_{b-1,(a-1)(b-1)}$ ونرمز لها بالرمز F_0' وهي:

$$F_0' = \frac{MSS_{blks}}{MSE}$$

حيث فترة قبول H_0 عند مستوى معنوية α هي $\left[0, F_{(b-1),(a-1)(b-1),\alpha}\right[$ ونوضح فترة قبول H_0 مع فترة رفض H_0 بالجزء المظلل في شكل (٢,١٢) كما يلي



شكل (٢,١٢)

ونرفض H_0 إذا كانت F_0 واقعة في منطقة رفض H_0 أي خارج فترة قبول H_0

أي

$$F_0' > F_{(b-1),\ (a-1)(b-1),\ \alpha}$$

ويمكن صياغة جدولا للحسابات يسمى تحليل التبايـن (ANOVA) كما في جدول (١٢,٣) كما يلي:

جدول (١٢,٣) تحليل التباين في اتجاهين

مصدر الاختلاف	مجموع المربعات SS	درجة الحرية d.f.	متوسط مجموع المربعات MSS	F_0
بين المعالجات	$SS_{trt}=\dfrac{\sum_{i=1}^{a} Y_{i.}^2}{b} - \dfrac{Y_{..}^2}{N}$	a - 1	$SS_{trt}/(a-1)$	$\dfrac{MSS_{trt}}{MSE}$
بين القطاعات	$SS_{blks}=\dfrac{\sum_{j=1}^{b} Y_{.j}^2}{a} - \dfrac{Y_{..}^2}{N}$	b - 1	$SS_{blks}/(b-1)$	$\dfrac{MSS_{blks}}{MSE}$
الخطأ	$SSE=SS_{Tot}-SS_{trt}-SS_{blks}$	(a-1)(b-1)	$SSE/(a-1)(b-1)$	
الكلي	$SS_{Tot}=\sum_{i}^{a}\sum_{j}^{b} y_{ij}^2 - \dfrac{Y_{..}^2}{N}$	N - 1		

حيث يحسب مجموع مربعات الخطـأ SSE بالطرح من العلاقة (١٢,٣) السابقة كما يلي:

$$SSE = SS_{Tot} - SS_{trt} - SS_{blks} \qquad (١٢,٧)$$

مثال (١ , ١٢)

إذا كان لدينا أربعة أنواع من السمــاد A, B, C, D استخدمت معالجات فــي تصميم القطاعات العشوائية الكاملة لمعرفة ما إذا كان لها تأثير مختلف على زيادة إنتاج القمح وكانت البيانات موضحة في الجدول (٤ , ١٢) التالي:

جدول (٤ , ١٢)

المعالجات	القطاع الأول	القطاع الثاني	القطاع الثالث	القطاع الرابع	المجموع $Y_{i.}$
A	9.3	9.4	9.6	10	38.3
B	9.4	9.3	9.8	9.9	38.4
C	9.2	9.4	9.5	9.7	37.8
D	9.7	9.6	10	10.2	39.5
المجموع Y_j	37.6	37.7	38.9	39.8	154.0

نجري الحسابات للقيم الموضحة في جدول (٣ , ١٢) لتحليل التبـايـن فـي اتجاهين كما يلي:

$$SS_{Tot} = \sum_{i}^{a}\sum_{j}^{b} y_{ij}^{2} - \frac{Y_{..}^{2}}{N}$$

$$= (9.3)^{2} + \dots + (10.2)^{2} - \frac{(154)^{2}}{16}$$

$$= 1483.54 - 1482.25 = 1.29$$

$$SS_{trt} = \frac{\sum_{i=1}^{a} Y_{i.}^{2}}{b} - \frac{Y_{..}^{2}}{N}$$

$$SS_{trt} = \frac{(38.3)^{2}}{4} + \frac{(38.4)^{2}}{4} + \frac{(37.8)^{2}}{4} + \frac{(39.5)^{2}}{4} - \frac{(154)^{2}}{16}$$

$$= 0.3850$$

$$SS_{blks} = \frac{\sum_{j=1}^{b} Y_{.j}^{2}}{a} - \frac{Y_{..}^{2}}{N}$$

$$SS_{blks} = \frac{(37.6)^{2}}{4} + \frac{(37.7)^{2}}{4} + \frac{(38.9)^{2}}{4} + \frac{(39.8)^{2}}{4} - \frac{(154)^{2}}{16}$$

$$= 0.825$$

ويحسب SSE من العلاقة (١٢,٧) كما يلي:

$$SSE = 1.29 - 0.385 - 0.825 = 0.08$$

يمكن وضع النتائج السابقة في جدول (١٢,٥) لتحليل التباين في اتجاهين كما يلي:

جدول (١٢,٥) تحليل التباين في اتجاهين

مصدر الاختلاف	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	F_0
	SS	d.f	MSS	
بين المعالجات	0.385	3	0.1283	$\frac{0.1283}{0.0089} = 14.42$
بين القطاعات	0.825	3	0.2750	$\frac{0.275}{0.0089} = 30.9$
الخطــــأ	0.08	9	0.0089	
الكلـــــي	1.29	15		

من الجدول (١٢,٥) السابق يمكن اختبار الفرض العـدمـي H_0 بين المعالجات ثم اختبار الفرض العدمي H_0 بين القطاعات كما يلي:

أولا: اختبار الفرض العدمي H_0 بين المعالجات يصاغ كما يلي:

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$

ضد الفرض البديل

$H_1 : \tau_i \neq 0$ واحدة على الأقل بحيث i

تحت صحة H_0 فإن منطقة قبول H_0 هي$(0,\ F_{3,9,0.05})$ وخارج هذه الفتـرة يعتبر رفضH_0 ومن جداول F (٤) في آخر الكتاب نجد أن هذه الفترة هي $(0, 3.86)$ من جدول (١٢,٥) السابق نجد$F_0 = 14.42$ واقعة خارج فترة قبـولH_0 أي أن نرفضH_0 أي المعالجات (السماد) لها تأثير مختلف في إنتاج القمح عند مستوى معنوية $\alpha = 0.05$.

ثانيا: اختبار الفرض العدميH_0 بين القطاعات يصاغ كما يلي:

$$H_0 : B_1 = B_2 = B_3 = B_4 = 0$$

ضد الفرض البديل

$H_0 : B_j \neq 0$ واحدة على الأقل بحيث j

تحت صحة H_0 فإن منطقة قبول الإحصاءة F هي$(0, F_{3,9,0.06})$ وخارج هـذه الفترة تعتبر منطقـة رفضH_0 ومن الجـداول نجـد أن منطقة قبـولH_0 هي $(0, 3.86)$ عند مستوى معـنـويـة $\alpha = 0.05$ ومن جدول تحليل التـبـايـن (١٢,٥) نجد$F_0 = 30.9$ واقعة خارج منطقة قبـولH_0 أي أننا نرفـضH_0 عند مستـوى معنوية $\alpha = 0.05$ وهذا يعني أن التأثير بين القطاعات معنوي وأن التحليل باستخدام تصميم القطاعات العشوائية الكاملة أفضل من التحليل تصميم التام التعشية (تحليل التباين في اتجاه واحد). ويمكن استنتاج جدول (١٢,٦) لتحليل التباين في اتجـاه واحد بإضافة مجموع مربعات بين القطاعات SS_{blks} إلى مجموع مربعات الخطـأ SSE وذلك باستخدام النتائج التي تم حسابها بجدول (١٢,٥) ونوضح ذلك في جدول (١٢,٦) كما يلي:

جدول (٦ , ١٢). جدول تحليل التباين في اتجاه واحد

مصدر الاختلاف	مجموع المربعات SS	درجات الحرية d.f	متوسط مجموع MSS	F_0
بين المعالجات	0.385	3	0.1283	1.7
الخطأ	0.905	12	0.0754	
الكلي	1.29	15		

وباختبار الفرض العدمي H_0 بين المعالجات حيث

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$

ضد الفرض البديل

$H_1 : \tau_i \neq 0$ واحدة على الأقل بحيث i

تحت صحة H_0 فإن منطقة قبول H_0 هي $(0, F_{3, 12, 0.05})$ ومنطقة رفض H_0 خارج هذه الفترة من جداول F (٤) آخر الكتاب نجد فترة قبول H_0 هي (0, 3.49) ومن جدول (٦ , ١٢) نجد أن F_0 المحسوبة هي 1.7 واقعة داخل منطقة قبول H_0 أي أن لا تأثير معنوي بين المعالجات (السماد) أي أن لها تأثير متساوي في إنتاج القمح وهذه النتيجة مختلفة عما تم في تحليل التباين (٥ , ١٢) في اتجاهين ما يوضح دقة التحليل في اتجاهين في هذا المثال عن تحليل التباين في اتجاه واحد.

(٤ , ١٢) المقارنات المتعددة

Multiple Comparison Procedures

إذا كانت الفروق بين متوسطات المعالجات معنوية فإننا نهتم غالبا بالمقارنات المتعددة لكي نعرف أي المعالجات التي سببت الفروق. نستخدم أي من الأساليب التي تم شرحها في الفصل الحادي عشر السابق والخاص بالمقارنات المتعددة. حيث

يتم ببساطة في أي صيغة من الصيغ نضع عدد القطاعات b بدلا من عدد التكرارات n واستخدام درجات الحرية الجديدة (a-1)(b-1) بدلا من n(a-1). فمثلا إذا أردنا استخدام اختبار دنكان فإن الخطأ المعياري هو $S_{\overline{Y}_i} = \sqrt{\frac{MSE}{b}}$ ودرجات حرية هي (a-1)(b-1) ونوضح ذلك بالمثال التالي:

مثال (١٢,٢)

في مثال (١٢,١) الخاص بالأسمدة وتأثيرها على إنتاج القمح حيث كانت الفروق بين المعالجات (السماد) معنوية وكان متوسطات المعالجات مرتبة ترتيبا تصاعديا كما يلي:

$$\overline{Y}_C = 9.45 \ , \quad \overline{Y}_A = 9.575 \ , \quad \overline{Y}_B = 9.6 \ , \quad \overline{Y}_D = 9.875$$

وقيمة الخطأ المعياري للمتوسطات $S_{\overline{Y}_i}$ هو.

$$S_{\overline{Y}_i} = \sqrt{\frac{MSE}{b}} = \sqrt{\frac{0.0089}{4}} = 0.047$$

من جداول دنكان رقم (٦) آخر الكتاب تحسب الأحصاءة R_p من العلاقة (١٢,٨) التالية:

$$R_\rho = r_\alpha(\rho, f)\, S_{\overline{Y}_{i.}} \qquad (١٢,٨)$$

حيث f درجات حرية الخطأ وفي مثالنا f = 12 و ρ رقم مستوى المعالجة وفي مثالنا ρ = 2,3,4 و R_p تمثل أقصر مدى للمعنوية. ويتم الحسابات من جدول (١٢,٧) كما يلي وذلك باعطاء المعالجات A,B,C,D الرتب 1,2,3,4 على الترتيب.

جدول (١٢,٧)

متوسط المعالجات مرتبة	$\overline{Y}_A$	$\overline{Y}_B$	$\overline{Y}_C$	$\overline{Y}_D$
متوسط المعالجات مرتبة حسب قيمها	$\overline{Y}_{1.}$	$\overline{Y}_{2.}$	$\overline{Y}_{3.}$	$\overline{Y}_{4.}$
قيم المتوسطات $\overline{Y}_{i.}$	9.45	9.575	9.6	9.875

من جدول دنكان (٦) آخر الكتاب نحسب $r_{\alpha}(\rho, f)$ كما يلي:

ρ	2	3	4
$r_{0.05}(\rho, 12)$	3.08	3.23	3.33

ويتم وضع قيم R_ρ والمعطاة بالعلاقة (١٢,٨) كما يلي:

ρ	2	3	4
R_ρ	0.145	0.152	0.157

ويمكن كتابة فروق المتوسطات $\overline{Y}_{i.}$ ومقارنتها بقيم R_p كما يلي

$$\overline{Y}_{4.} - \overline{Y}_{3.} = 9.875 - 9.45 = 0.425 > 0.157 \qquad (R_4)$$
$$\overline{Y}_{4.} - \overline{Y}_{1.} = 9.875 - 9.575 = 0.3 > 0.157 \qquad (R_3)$$
$$\overline{Y}_{4.} - \overline{Y}_{2.} = 9.875 - 9.6 = 0.275 > 0.152 \qquad (R_2)$$
$$\overline{Y}_{2.} - \overline{Y}_{3.} = 9.6 - 9.45 = 0.15 < 0.152 \qquad (R_3)$$
$$\overline{Y}_{2.} - \overline{Y}_{1.} = 9.6 - 9.57 = 0.025 < 0.145 \qquad (R_2)$$
$$\overline{Y}_{1.} - \overline{Y}_{3.} = 9.575 - 9.45 = 0.135 < 0.145 \qquad (R_2)$$

نلاحظ من المقارنات السابقة أننا لا نستطيع رفض H_0 عندما

$$H_0: \mu_2 = \mu_3 \quad \text{أي} \quad (H_0: \mu_B = \mu_C)$$

أو

$$H_0: \mu_2 = \mu_1 \quad \text{أي} \quad (H_0: \mu_B = \mu_A)$$

أو

$$H_0: \mu_1 = \mu_3 \quad \text{أي} \quad (H_0: \mu_A = \mu_C)$$

لكن الرفض يتم عندما:

$$H_0: \mu_1 = \mu_4 \quad \text{أي} \quad (H_0: \mu_A = \mu_D)$$

$$H_0: \mu_2 = \mu_4 \quad \text{أي} \quad (H_0: \mu_B = \mu_D)$$

$$H_0: \mu_3 = \mu_4 \quad \text{أي} \quad (H_0: \mu_C = \mu_D)$$

نستنتج مما سبق أن المتوسطات الثلاثة μ_A, μ_B, μ_C غير معنوية بينما المتوسط الرابع μ_D معنويا مع جميع المتوسطات الآخرى μ_A, μ_B, μ_C.

يريد البعض في بعض الأحيان تجزىء مجموع مربعات المعالجات SS_{trt} إلى مجاميع مربعات خاصة بمقارنات متعامدة بحيث يكون مجاميع المربعات خاصة بالمقارنات مساويا لمجموع مربعات المعالجات ونوضح ذلك بالمثال التالي.

مثال (١٢,٣)

افترض لدينا تجربة صممت طبقا لتصميم القطاعات العشوائية الكاملة. تم الحصول على النتائج وتحليلها وحصلنا على جدول تحليل التباين (١٢,٨) كما يلي:

جدول (١٢,٨)

مصدر الاختلاف	مجموع المربعات SS	درجات الحرية d.f	متوسط مجموع MSS	F_0
بين المعالجات	1210.58	7	172.94	29.95**
بين القطاعات	102.3	5	20.5	
الخطأ	202.17	35	5.87	

حيث كان مجموع كل معالجة كما هو موضح في جدول (١٢,٩) كما يلي:

جدول (١٢,٩)

رقم المعالجة	1	2	3	4	5	6	7	8
مجموع مشاهدات المعالجة	44	119	84	51	22	32	33	65

لدينا هنا 7 درجات حرية للمعالجات أي أننا نريد تكوين 7 مقارنات متعامدة لكي نجزىء مجموع مربعات المعالجات. أي درجة حرية واحدة لكل مقارنة والفروض المراد دراستها توضح كما يلي:

$$H_0: \mu_1 = \mu_2 + \mu_3 + .. + \mu_7 + \mu_8$$

أي أن $-7Y_{1.} + Y_{2.} + Y_{3.} + Y_{4.} + Y_{5.} + Y_{6.} + Y_{7.} + Y_{8.} = 0$

$$H_0: \mu_2 + \mu_3 = \mu_4 + \mu_5 + \mu_6 + \mu_7 + \mu_8$$

$$5Y_{2.} + 5Y_{3.} - 2Y_{4.} - 2Y_{5.} - 2Y_{6.} - 2Y_{7.} - 2Y_{8.} = 0$$

$$H_0: \mu_2 = \mu_3$$

$$Y_{2.} - Y_{3.} = 0$$

$$H_0: \mu_4 + \mu_8 = \mu_5 + \mu_6 + \mu_7$$

$$3Y_{4.} - 2Y_{5.} - 2Y_{6.} - 2Y_{7.} + Y_{8.} = 0$$

$$H_0 : \mu_4 = \mu_8$$

$$Y_{4.} \qquad -Y_{8.} = 0$$

$$H_0 : \mu_5 = \mu_6 + \mu_7$$

$$2Y_{5.} - Y_{6.} - Y_{7.} = 0$$

$$H_0 : \mu_6 = \mu_7$$

$$Y_{6.} - Y_{7.} = 0$$

وباستخدام بيانات مثال (٣,١٢) السابق وحساب مجاميع المربعات المتعامدة SSc للمقارنة يمكننا تلخيص المقارنات في جدول (١٢,١٠) كما يلي:

جدول (١٢,١٠)

مجموع المربعات	$Y_{1.}$	$Y_{2.}$	$Y_{3.}$	$Y_{4.}$	$Y_{5.}$	$Y_{6.}$	$Y_{7.}$	$Y_{8.}$	I	II	SSc
للمعالجات	**44**	**119**	**84**	**51**	**22**	**32**	**33**	**65**	$\Sigma c_i y_i$	$b\Sigma C_i^2$	I^2/II
C_1	-7	1	1	1	1	1	1	1	98	6(56)	28.583
C_2	0	5	5	-2	-2	-2	-2	-2	609	6(70)	888.05
C_3	0	1	-1	0	0	0	0	0	35	6(2)	102.083
C_4	0	0	0	+3	-2	-2	-2	3	174	6(30)	168.20
C_5	0	0	0	1	0	0	0	-1	-14	6(2)	16.33
C_6	0	0	0	0	2	-1	-1	0	-21	6(6)	12.25
C_7	0	0	0	0	0	1	-1	0	-1	6(2)	0.083
										SS_{trt}	1210.58

ولبيان كيفية الحساب في جدول (١٢,١٠) للمجاميع $Y_{i.}$ $\sum_{i=1}^{\infty} c_i Y_{i.}$ و $b\Sigma c_i^2$

وكذلك مجموع مربعات المقارنات المتعامدة SSc نوضح ذلك من المقارنة الأولى C_1 كما يلي:

$$I = \sum_{i=1}^{8} c_i Y_{i.} = (-7)44 + (1)119 + (1)51 + (1)22 + (1)32 + (1)3\!:$$

$$+ (1)65 = 98$$

وكذلك $II = b\Sigma c_i^2 = 6(49+1+1+1+1+1+1+1) = 5$

$$SSc_1 = \frac{I^2}{II} = (98)^2/56 = 28.583$$

وهكذا لباقي المقارنات $C_2, C_3, ..., C_7$

ويمكن صياغة جدول تحليل التباين (١١ , ١٢) ممثلا للجدول (٨ , ١٢) السابق مع إضافة المقارنات المتعامدة كما يلي:

جدول (١١ , ١٢) تحليل التباين للمقارنات المتعامدة (ANOVA)

مصدر الاختلاف	مجموع المربعات SS	درجات الحرية d.f	متوسط مجموع المربعات MSS	F_0
بين المعالجات	1210.58	7	172.94	29.95
C_1	28.583	1	28.585	4.95
C_2	888.05	1	888.05	153.64
C_3	102.083	1	102.083	17.79
C_4	168.20	1	168.20	29.1
C_5	16.33	1	16.33	2.83
C_6	12.25	1	12.25	2.12
C_7	.08	1	0.08	0.014
بين القطاعات	102.3	5	20.5	3.55
الكلـــي	202.17	35	5.78	--

نلاحظ من جدول (١٠ , ١٢) أن مجموع مربعات بين المعالجات يساوي مجموع مربعات تلك المقارنات المتعامدة أي أن $(\sum_{i=1} SSc_i = 1210.58)$ وأن مقارنة F_0 بقيمة F من الجدول وهي $F_{1,35,0.05} = 4.17$ نلاحظ معنوية المقارنات C_1, C_2, C_3, C_4 وأن المقارنات C_5, C_6, C_7 غير معنوية لأن F_0 لها أقل من $F_{1,35,0.05} = 4.17$ الجدولية.

(٥ , ١٢) تمارين

Exercises

١ - قام أحد الباحثين في علم النفس بدراسة الانفعالات على القدرة الكامنة الكهربائية للجلد بالمليميكوفولت لمجموعة من الأفراد تحت انفعالات مختلفة وهي الخوف والسعادة والاكتئاب والهدوء وحصل على البيانات في الجدول التالي:

الشخص	الخوف	السعادة	الاكتئاب	الهدوء
1	23.1	22.7	22.3	22.6
2	57.6	53.2	53.7	23.1
3	10.5	09.7	10.8	08.3
4	23.6	19.6	21.1	21.0
5	11.9	13.8	13.2	13.3
6	54.0	47.1	39.2	37

(i) حلل البيانات التي بالجدول السابق.

(ii) افترض أن الباحث يرغب في مقارنة الاكتئاب ضد بقية الانفعالات. ضع المقارنة المناسبة واختبر معنويتها.

٢ -يرغب أحد الباحثين في مقارنة ثلاثة أنواع من المحاليل على نمو البكتريا في وعاء (يحتوي على 5 جالونات). تم التحليل في المعمل وكان لا يمكن إجراء التجربة إلا ثلاث مرات في اليوم. ويعتقد الباحث أن اختلاف الأيام يؤدي إلى اختلاف ملحوظ فور استخدام القطاعات العشوائية الكاملة معتبرا الأيام قطاعات، وحصل على النتائج الموضحة في الجدول التالي:

القطاعات / المعالجات	اليوم الأول	اليوم الثاني	اليوم الثالث	اليوم الرابع
المحلول الأول	13	22	16	39
المحلول الثاني	16	24	17	44
المحلول الثالث	5	4	1	22

(i) حلل البيانات التي في الجدول السابق وناقش النتائج.

(ii) جزء مجموع المربعات للمعالجة إلى ثلاثة مجاميع لمقارنات متعامدة.

٣ - أجرى أحد الباحثين تجربة على طول الوقت الذي يمكن أن تركز فيه العين على أحد الأهداف، وكان يعتقد أن المسافة والهدف يؤثران على زمن التركيز، وكان يهتم بأربعة أزمنة، وكان لديك خمسة أشخاص لإجراء التجربة، وكان يعتقد أيضا أن هناك فرق بين الأفراد. فقرر إجراء تجربة استخدم فيها تصميم القطاعات العشوائية الكاملة مستخدما الأفراد كقطاعات. حيث كانت النتائج كما هو موضح بالجدول التالي:

القطاعات / المعالجة	الشخص الأول	الشخص الثاني	الشخص الثالث	الشخص الرابع	الشخص الخامس
المسافة 4 قدم	10	6	6	6	6
المسافة 6 قدم	7	6	6	1	6
السافة 8 قدم	5	3	3	2	5
المسافة 10 قدم	6	4	4	2	3

أجر تحليل التباين للنتائج بالجدول السابق وناقش النتائج.

٤ - أراد أحد الكيميائيين أن يختبر تأثير 4 محاليل كيميائية على قوة نوع من القماش، وكان يعتقد أن هناك اختلافا من ثوب لآخر. فقرر اعتبار الثوب قطاعا، واجرى التجربة وحصل على النتائج كما في الجدول التالي:

القطاعات / المعالجات	الثوب الأول	الثوب الثاني	الثوب الثالث	الثوب الرابع	الثوب الخامس
المحلول الأول	73	68	74	71	67
المحلول الثاني	73	67	75	72	70
المحلول الثالث	75	68	78	73	68
المحلول الرابع	73	71	75	75	79

أجر تحليل التباين للنتائج بالجدول السابق وناقش النتائج.

٥ - يرغب أحد الباحثين الزراعيين في دراسة إنتاجية أربعة أنواع من القمح. وقام بتقسيم الحقل إلى ستة قطاعات، وقام بتوزيع أنواع القمح عشوائيا داخل كل قطاع (كل قطاع مكون من أربعة قطع)، وحصل على النتائج كما هو موضح بالجدول التالي (المحصول بالرطل لكل قطعة).

المعالجات / القطاعات	أنواع القمح			
	A	B	C	D
1	30.4	28.8	33.0	31.8
2	33.9	25.5	32.7	33.5
3	32.7	27.3	34.5	34.5
4	34.9	29.3	36.0	39.5
5	31.9	27.5	36.5	34.5
6	35.1	28.3	34.2	36.0

(i) هل هناك فرق معنوي بين المحصول لأنواع القمح المختلفة $\alpha = 0.05$.

(ii) إذا رغب الباحث في مقارنة النوع B ضد بقية الأنواع ضع المقارنة المتعددة المناسبة واختبر معنويتها.

(iii) جزء مجموع المربعات للمعالجات إلى ثلاثة مجاميع لمقارنات متعامدة على أن يكون من بينها (B).

٦ - كوّن جدول تحليل التباين التالي ثم:

(i) اختبر معنوية المعالجات وأيضا معنوية القطاعات.

(ii) افترض أن التجربة تم تحليلها مع اهمال القطاعات (أي في اتجاه واحد) ثم اكتب جدول تحليل التباين في هذه الحالة وأجر الاختبارات المناسبة وعلق على النتائج.

المصدر	مجموع المربعات	درجات الحرّية	متوسط مجموع المربعات	F_0
بين المعالجات		2		10
بين القطاعات				
للخطأ		6	4	
الكلي	120	11		

الفصل الثالث عشر

الارتباط والانحدار
Correlation and Regression

● مقدمة ● المخطط الانتشاري ● معامل الارتباط الخطي لبيرسون ● اختبارات الفروض حول معامل الارتباط الخطي ● معامل الارتباط للرتب لسبيرمان ● اختبارات الفروض حول معامل الارتباط للرتب ● معامل الاقتران ومعامل التوافق ● الانحدار الخطي البسيط ● اختبارات الفروض حول معامل الانحدار وتقدير فترة الثقة له ● معامل التحديد ● معادلات انحدار غير خطية ويمكن تحويلها إلى معادلات خطية ● الانحدار الخطي المتعدد ● تمارين.

(١٣,١) مقدمة
Introduction

لقد درسنا حتى الآن المتغير العشوائي لمجموعة واحدة من المشاهدات وذلك في الفصول السابقة. في هذا الفصل سوف ننظر إلى مجموعتين مختلفتين من المشاهدات ونرى فيما إذا كانت هناك علاقة ما بينهما، وهل التغير في إحدى هذه المجموعات يصاحبه تغير في المجموعة الأخرى، وفي أي اتجاه وبأي كمية. ففي مجال الإحصاء توجد عدة مشاكل تحتاج إلى تحليل أكثر من متغير واحد. فعلى سبيل المثال في مجال الأعمال ومجال الاقتصاد ومجال التعليم ومجال الصحة وغيرها كثيرا نحاول الإجابة على أسئلة مثل: هل هذان المتغيران توجد بينهما علاقة ما؟ وإذا كانت بينهما علاقة فما هي؟ هل هذه المتغيرات مرتبطة؟ سوف

نقوم ببناء علاقات رياضية تتنبأ بتصرف متغير ما في معرفة التغير في متغير آخر، فمثلا قد نتساءل هل هناك علاقة بين الطول والوزن لمجموعة من الأشخاص؟ هل هناك علاقة للتعرض للإشعاع والإصابة بمرض السرطان؟ هل هناك علاقة بين عدد ساعات التجول بسيارة ما في شوارع مدينة الرياض واحتمال الإصابة في حادث مروري؟ هل هناك علاقة بين عدد ساعات المذاكرة لمادة ما والدرجات التي يحصل عليها الطالب في الامتحان؟ هل هناك علاقة بين كمية الدواء وسرعة شفاء المريض؟ هل هناك علاقة بين النشاط الإعلاني وحجم المبيعات لمؤسسة ما؟... الخ.

وللإجابة على هذه الأسئلة نحتاج لدراسة العلاقة بين مجموعتين من القراءات، نحتاج أولا إلى ترتيبها على شكل أزواج مرتبة (x, y) حيث x هي قراءة المتغير الأول (في المجموعة الأولى) و y هي قراءة المتغير الثاني (في المجموعة الثانية)، والسبب في تسميتها أزواج مرتبة هو أن x قراءة المتغير الأول تكتب دائما أولا وتتبعها y قراءة المتغير الثاني وهما يشكلان زوجا من القراءات والتي نحافظ على ترتيبها في جميع الوحدات المشاهدة. فمثلا في أحد الأمثلة السابقة إذا كان لدينا مجموعة من الأشخاص فإن x سوف تكون طول الشخص و y وزن الشخص ونسجل هذه للشخص الأول على شكل (x_1, y_1) وللشخص الثاني (x_2, y_2) وهكذا حتى الشخص ذو الرتبة n (x_n, y_n)، وأيضا تسمى x قراءة المتغير المدخل (input variable) أو المتغير المستقل (independent variable)، y قراءة المتغير المخرج output variable أو قراءة المتغير التابع (dependent variable). قراءة المتغير x أو المتغير المدخل أو المستقل أحيانا يكون مقاس أو مسيطر عليه أو متحكم فيه بحيث يكون الغرض هو التنبؤ عن قراءة المتغير y المتغير المخرج أو التابع. فمثلا في مثال الدواء يمكن التحكم في كمية الدواء وتعيينها مسبقا ونرمز لها بالرمز x (في هذه الحالة x يكون غير عشوائي) وفي مثال الطول والوزن يمكن أن يكون أيا من الطول x والوزن y المتغير المستقل ويكون الآخر المتغير التابع (في هذه الحالة يكون كلاهما عشوائي).

(١٣,٢) المخطط الانتشاري

Scatter Diagram

قبل إجراء تحليل رياضي على البيانـات المشـاهدة لغرض معرفة فيمـا إذا كـان هناك أي علاقة بين أي متغيرين، فإنه من الأفضل رسم هذه المشـاهدات فيـما يسمى بالمخطـط الانتشـاري (Scatter diagram)، وفيه نرسم المتغير x يتبع إحداثيات المحور الأفقي السيني والمتغيـر y والذي يتبع احداثيات المحور الرأسي الصـادي، وهذه المخططات تعطي صورة أولية تساعد كثيرا في أخذ فكرة أولية جيـدة عـن العلاقة بين المتغيرين، وسوف نوضح ذلك بالمثال التالي.

مثال (١٣,١)

البيانات التالية والتي هي على شكل زوج مرتب، هي عدد الساعات x التي يقضيها الطالب في استذكار مادة ما والدرجـات y التي حصل عليها في امتحان تلك المادة كما يلي:

(2,3),(3,5),(3,7),(4,5),(4,7),(5,7),(5,8),(6,6), (6,9), (7,7), (7,9),(7,10), (8,8), (8,9).

ارسم مخطط انتشار لهذه البيانات.

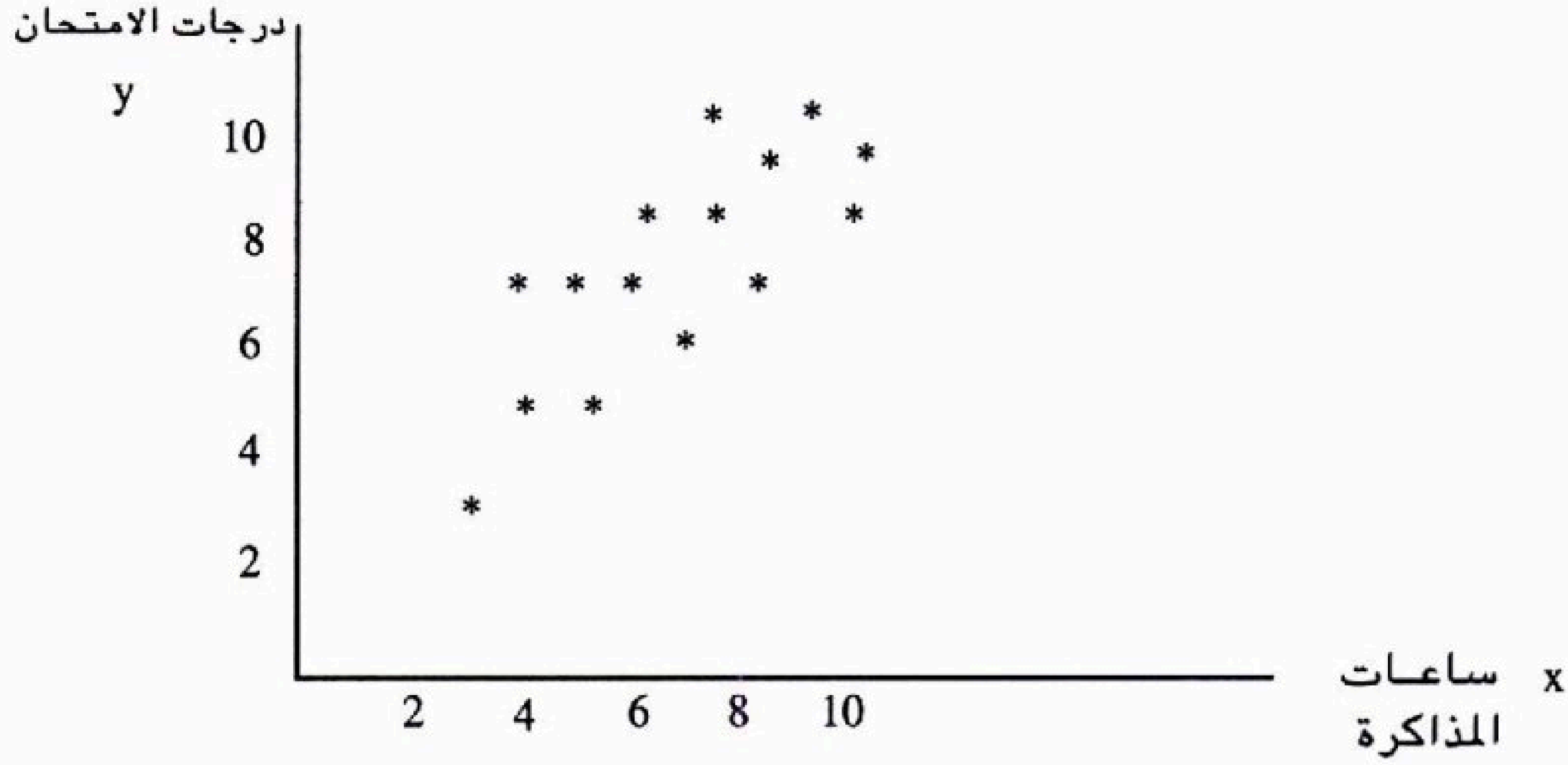
الحل

نضع بيانات الأزواج التي بالمثال في جدول (١٣,١) التالي:

جدول (١٣,١)

ساعات الاستذكار x	2	3	3	4	4	5	5	6	6	7	7	7	8	8
درجة الامتحان y	3	5	7	5	7	7	8	6	9	7	9	10	8	9

ثم نرسم بيانات جدول (١٣,١) في مخطط انتشار شكل (١٣,١) التالي:



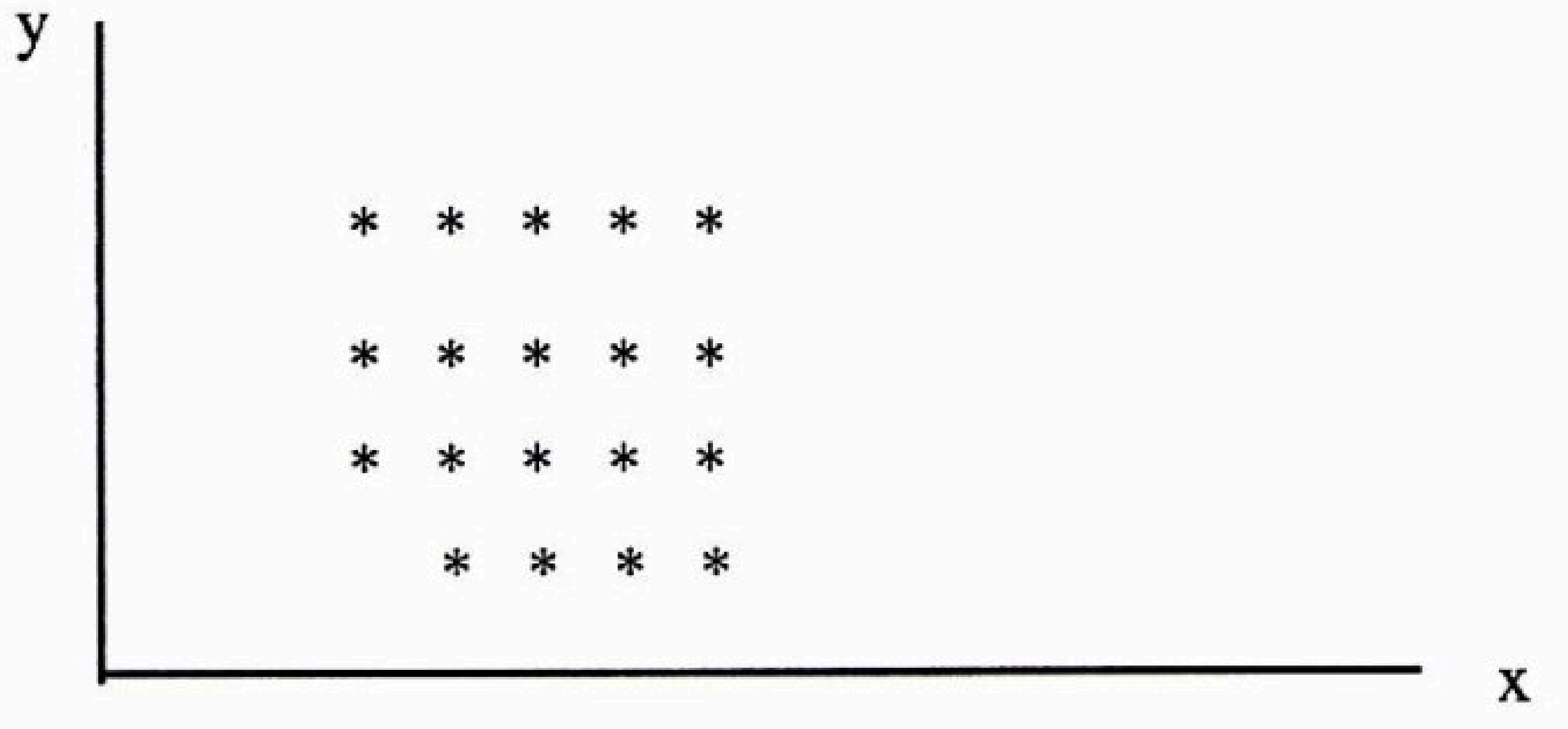
شكل (١ , ١٣)

نلاحظ من شكل الانتشار (١ , ١٣) أن هناك علاقة طردية بين عدد ساعات المذاكرة والدرجات التي حصل عليها الطالب في الامتحان، وهذه العلاقة غير كاملة بمعنى أن هناك تغيرعشوائي ضمن مجموعة الطلبة الذين استذكروا بعدد الساعات نفسها مثل الطلاب الذين استذكروا 6 ساعات وعددهم ثلاثة كانت درجاتهم على التوالي 6,9,8 واختلاف الدرجات هذا عائد إلى عوامل أخرى مثل الذكاء والرغبة للمادة. . . الخ.

(١ , ٢ , ١٣) أشكال الانتشار والارتباط الخطي

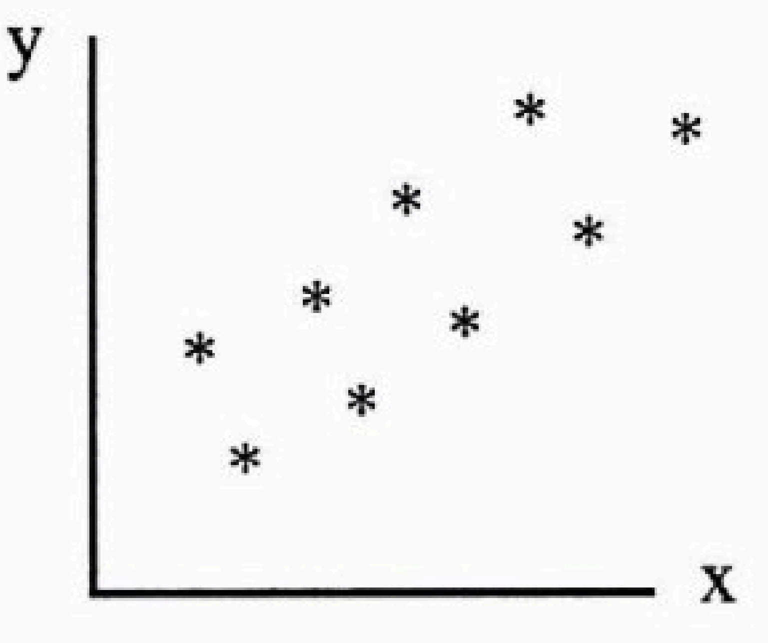
الغرض الأساسي من تحليل الارتباط الخطي هو قياس قوة العلاقة الخطية بين متغيرين. سوف نستعرض الآن فيما يلي بعض العلاقات الممكنة بين المتغير المستقل x والمتغير التابع y.

أولا: إذا زادت قيمة x فإنه لا يوجد تغير واضح في قيمة y، وفي هذه الحالة نقول x و y غير مرتبطة، أو لا يوجد علاقة بين x و y وهذه الحالة كما هو موضح في شكل (٢ , ١٣) التالي:

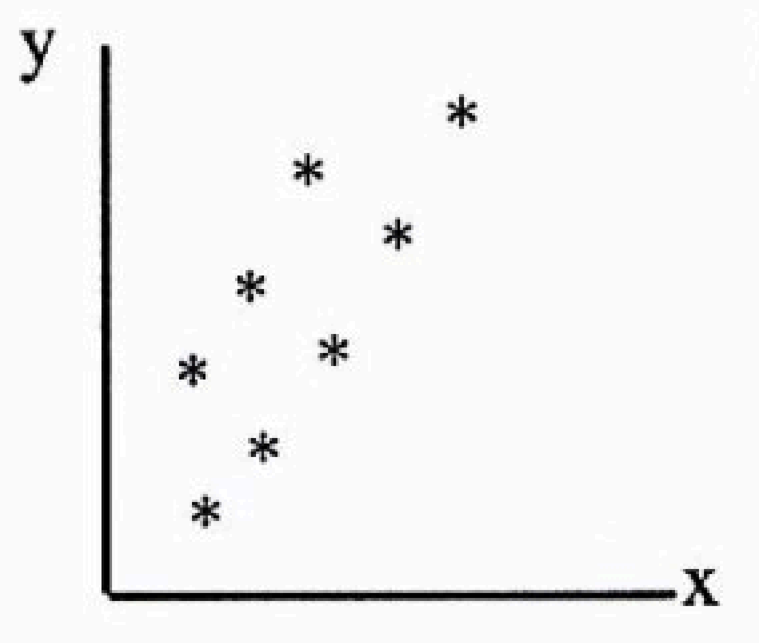


شكل (١٣,٢) لا يوجد إرتباط

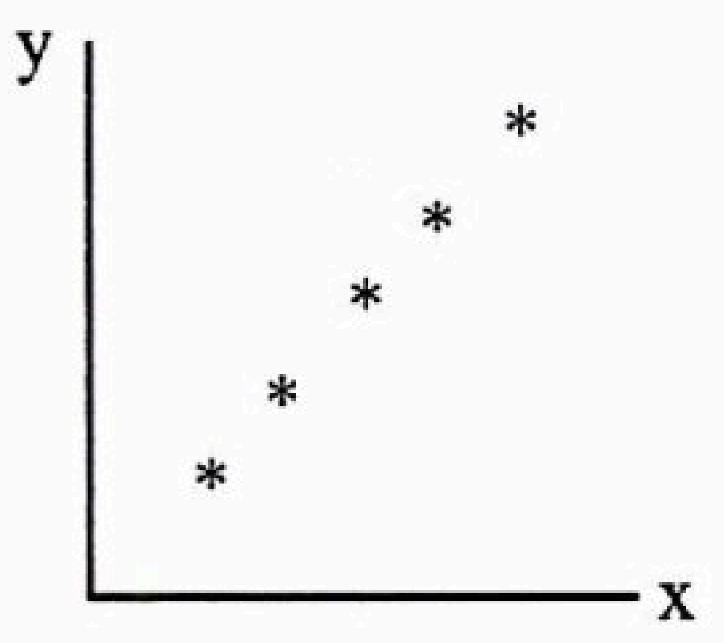
ثانيا: إذا زادت قيمة x فإن قيمة y تزداد بدرجات مختلفة، وفي هذه الحالة يكون هناك إرتباط موجب بين x و y، أو هناك علاقة موجبة بين x و y وقد تكون هذه العلاقة قوية جدا أو متوسطة أو ضعيفة، ويمثل ذلك بالأشكال (١٣,٣)، (١٣,٤) و (١٣,٥) التالية:



شكل (١٣,٥)
ارتباط موجب ضعيف



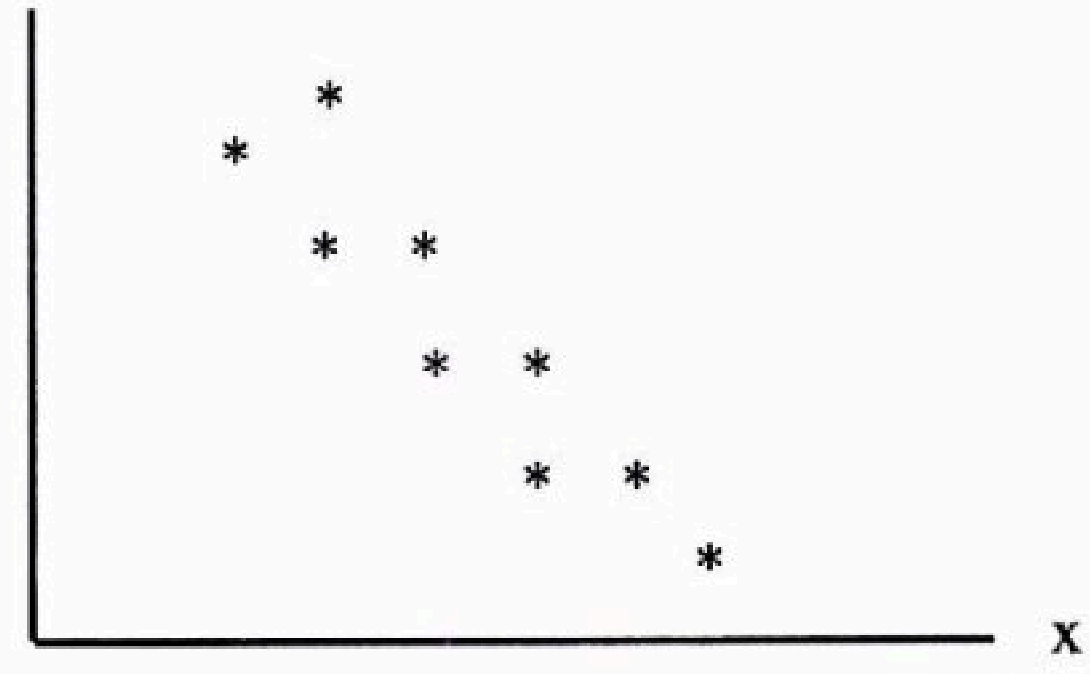
شكل (١٣,٤)
ارتباط موجب قوي



شكل (١٣,٣)
ارتباط كامل قوي

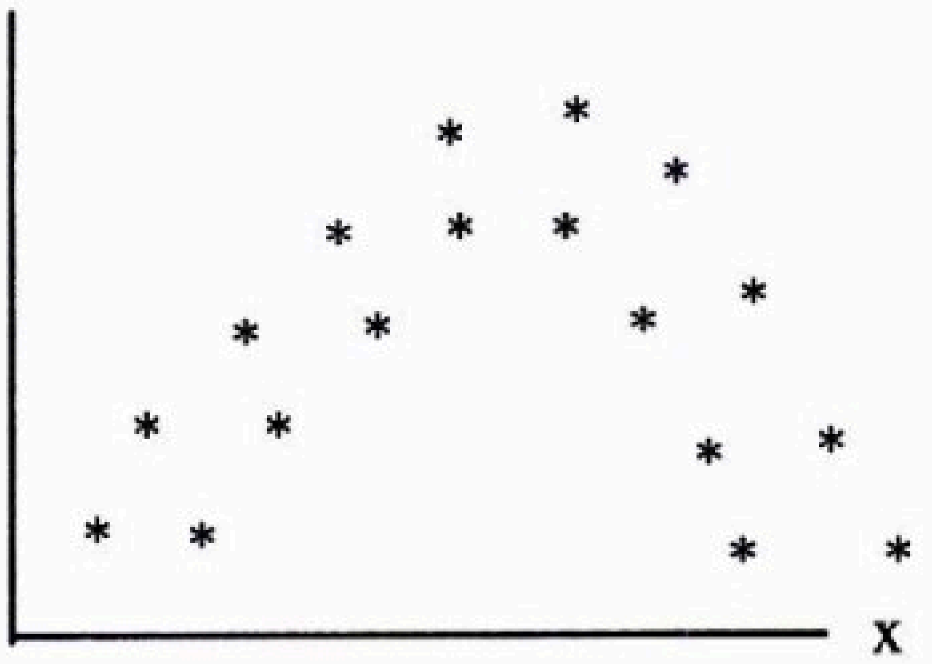
وشكل (١٣,٣) يمثل ارتباط خطي مثالي وذلك عندما تقع جميع النقــاط على خط مستقيم.

ثالثا: إذا زادت قيمة x فإن قيمة y تتناقص بدرجات مختلفة وفي هذه الحالة يكون هناك ارتباط سالب بين x و y أو علاقة سالبة بين x و y ويمثل ذلك بالشكل (٦, ١٣) كما يلي:



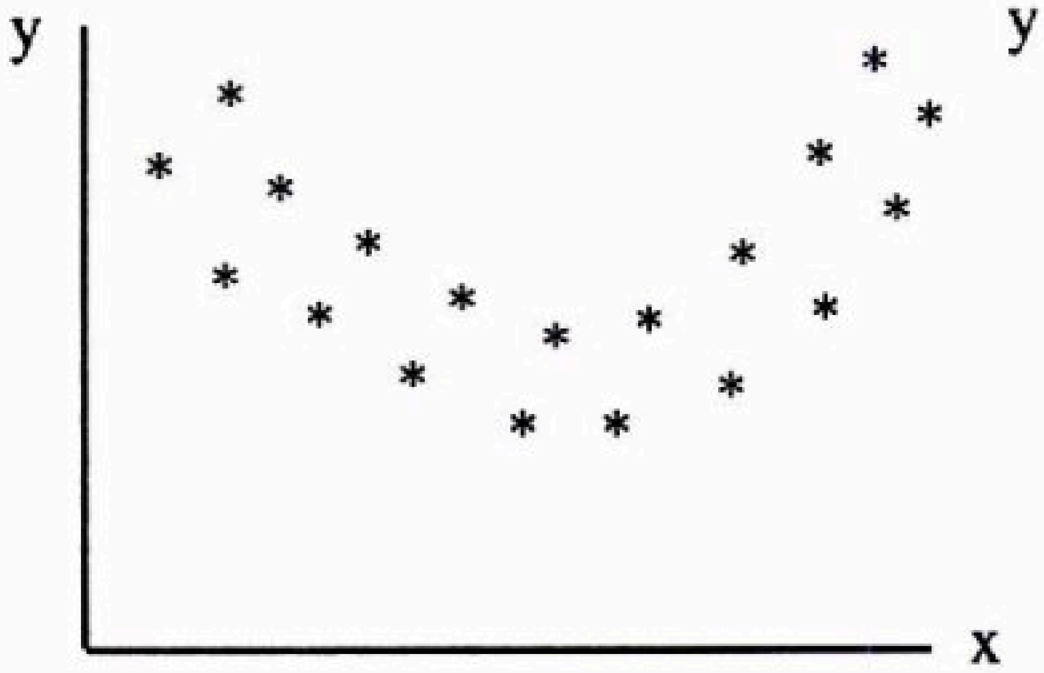
شكل (٦, ١٣) ارتباط سالب قوي

رابعا: قد يكون هناك نمط يمثل تغير y عندما تتغير x، لكن قد يحدث أن يكون تغيــر y عندما تتغير x غير خطي، أي أن تكون فيه النقاط منتشرة حول منحنى، أي توجد علاقة بين x و y غير خطية، ونوضح ذلك كما في الشكلين (٧, ١٣) و (٨, ١٣) كما يلي:



شكل (٨, ١٣)

ارتباط غير خطي



شكل (٧, ١٣)

ارتباط غير خطي

ملاحظة

في هذا الفصل سوف نركز فقط على الارتباط الخطي والعلاقات الخطيـة وكذلك على العلاقات التي يمكن تحويلها إلى علاقات خطية.

(١٣,٣) معامل الارتباط الخطي لبيرسون

Linear Correlation Coefficient of Person

معامل الارتباط الخطي ويرمز له بالرمز r هو عبارة عن مقياس لقوة العلاقة الخطية بين متغيرين. والمعامل يعكس مدى تماسك التأثير الناتج عن التغير في متغيـر x على التغير في المتغير الآخر y. ومعامل الارتباط الخطي تكون قيمتـه دائما بين 1 و 1-، وعندما تكون قيمته موجبة فإننا نقول إن العلاقة الخطية بين المتغيرين موجبة أو طردية، أما إذا كانت قيمته سالبة فإن العلاقة الخطية بين المتغيرين تكون سالبة أو عكسية. فمثلا الطول والوزن في الأطفال بينهما علاقة موجبة أو طرديـة لأن كلما ينمو الطفل ويزداد طوله يزداد وزنه، أما في حالة عمر سيارة وسعرها فإن العلاقة تكون سالبة أو عكسية فكلما زاد عمر السيارة قل سعرها.

ولمعامل بيرسون عدة صيغ رياضية متكافئة لحساب معامل الارتباط للبـيانات على شكل أزواج مرتـبـة $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ حيث معامـل الارتباط r لبيرسون يعطي بمتوسط حاصل ضرب القيم المعيارية للمتغيرين x و y أي أن:

$$r = \frac{1}{(n-1)} \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{s_X} \cdot \frac{(y_i - \bar{y})}{s_Y} \qquad (١٣,١)$$

حيث s_X, s_Y هي الانحراف المعياري للمتغيرن x و y على التوالي:

$$s_x^2 = \frac{1}{n-1}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right), \quad s_y^2 = \frac{1}{n-1}\left(\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right)$$

ويمكن كتابة العلاقة (١٣,١) السابقة بالصورة التالية:

$$r = \frac{SS(xy)}{\sqrt{SS(x)\,SS(y)}} \qquad (١٣,٢)$$

حيث

$$SS(x) = \Sigma x^2 - \frac{(\Sigma x)^2}{n}, \quad SS(y) = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$$

$$SS(xy) = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}$$

مثال (١٣,٢)

الجدول التالي يبين درجات مجموعة من 8 طلاب في كل من مادتي الإحصاء والرياضيات في أحد امتحانات الأعمال الفصلية كما يلي:

الإحصاء x	13	9	19	15	11	8	16	11
الرياضيات y	15	7	17	15	10	9	14	10

هل هناك علاقة بين تحصيل الطلاب في المادتين؟

الحل

لتبسيط الحل نكون جدول (١٣,٢) كما يلي:

جدول (١٣,٢)

x	y	$x-\bar{x}$	$y-\bar{y}$	x^2	y^2	xy	$(x-\bar{x})(x-\bar{y})$	
13	15	0.25	2.875	169	225	195	0.71875	
9	7	-3.75	-5.125	81	49	63	19.21875	
19	17	6.25	4.875	361	289	323	30.46875	
15	15	2.25	2.875	225	225	225	6.46875	
11	10	-1.75	-2.125	121	100	110	3.71875	
8	9	-4.75	-3.125	64	81	72	14.84375	
16	14	3.25	1.875	256	196	224	6.09375	
11	10	-1.75	-2.125	121	100	110	3.71875	
102	97	0	0	1398	1265	1322	85.25	المجموع

ومن جدول (١٣,٢) نحسب S_x و S_y كما يلي:

$$s_x^2 = \frac{1}{7}\left(1398 - \frac{(102)^2}{8}\right) = 13.928571$$

أي أن

$$s_x = \sqrt{13.928571} = 3.7321$$

وأن

$$s_y^2 = \frac{1}{7}\left[1265 - \frac{(97)^2}{8}\right] = 12.696428$$

أي أن

$$s_y = \sqrt{12.696438} = 3.5632047$$

من هذه المقادير السابقة لقيـم S_x و S_y و الجدول (١٣,٢) وباستخدام الصيغـة (١٣,١) نحسب معامل الارتباط r كما يلي:

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1)\, s_x\, s_y} = \frac{85.25}{7(3.7321)(3.5632047)} = \frac{85.25}{93.087652}$$

$$= 0.9158035$$

وباستخدام الصيغة (١٣,٢) نجد أن

$$SS(xy) = 1322 - \frac{(102)(97)}{8} = 85.25$$

$$SS(x) = 1398 - \frac{(102)^2}{8} = 97.5$$

$$SS(y) = 1265 - \frac{(97)^2}{8} = 88.875$$

$$r = \frac{SS(xy)}{\sqrt{SS(x).SS(y)}} = \frac{85.25}{\sqrt{(97.5)(88.875)}} = \frac{85.25}{93.08766} = 0.9158034$$

نلاحظ أن الصيغتين (١٣,١) و (١٣,٢) تعطي نفس النتيجة ولكن الصيغة الثانية أسهل بكثير في الحساب من الصيغة الأولى.

ملاحظة

هناك صيغة مبسطة للصيغة (١٣,١) وهي كما يلي:

$$r = \frac{n \Sigma x y - \Sigma x \Sigma y}{\sqrt{[n \Sigma x^2 - (\Sigma x)^2]\,[n \Sigma y^2 - (\Sigma y)^2]}} \qquad (١٣,٣)$$

وبالتعويض من جدول (١٣,٢) في الصيغة (١٣,٣) نحصل على معامل الارتباط r كما يلي:

$$r = \frac{8(1322) - (102)(97)}{\sqrt{[8(1398) - (102)^2][8(1265) - (97)^2]}} = \frac{682}{\sqrt{(780)(711)}}$$

$$= \frac{682}{744.70128} = 0.9158030$$

ملاحظة

هذه الصيغة (١٣,٣) والصيغة (١٣,١) أكثر عرضة للخطأ في الحسابات ولذلك ينصح باستخدام الصيغة (١٣,٢).

(١٣,٤) اختبار الفروض حول معامل الارتباط الخطي

Testing of Hypothesis for Linear Correlation Coefficient

الآن وبعد أن حسبنا معامل الارتباط الخطـي r في العينة المعطاة في مثـال (١٣,٢)، يخطر بذهننا السؤال التالي: هل قيمة r المحسوبة من الصيغ (١٣,١)، (١٣,٢) و (١٣,٣) السابقة تدل على أن هناك علاقة بين المتغيرين في المجتمـع الذي سحبت منه العينة محل الدراسة.

للاجابة على هذا السؤال نقوم باختبار فرض العدم H_0 كما يلي:

H_0: المتغيران غير مرتبطان خطيا

ضد الفرض البديل H_1 وهو فرض بديل ذو طرفين كما يلي:

H_1: المتغيران مرتبطان خطيا

في الحقيقة الفرض البديل H_1 يمكن أن ينقسم إلى قسمين حسب الوضع فمثـلا هناك فرض بديل H_1 بطرف واحد وهو أن نقول:

H_1: المتغيران مرتبطان خطيا بشكل موجب

أو

H_1: المتغيران مرتبطان خطيا بشكل سالب

وبشكل رياضي إذا كانت ρ ترمز لمعامل الارتباط الخطي للمجتمع فإن فرض العدم H_0 والفرض البديل H_1 يمكن وضعهما بالشكل التالي:

$$H_0 : \rho = 0$$

والفرض البديل H_1 ذو طرفين وهو:

$$H_1 : \rho \neq 0$$

والفرض البديل بطرف واحد وهو:

$$H_1 : \rho > 0 \quad \text{أو} \quad H_1 : \rho < 0$$

لاختبار الفروض السابقة H_0 و H_1 نحتاج إلى:

(i) حساب الإحصاءة t_0 لمعامل الارتباط r المحسوب من العينة حيث قيمة t_0 تحسب تحت صحة فرض العدم H_0 كما يلي:

$$t_0 = \frac{r - \rho}{s_r} = \frac{r - 0}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{r\sqrt{n-2}}{\sqrt{1 - r^2}} \qquad (١٣,٤)$$

ولها توزيع t_υ بدرجات حرية $\upsilon = n - 2$

حيث الخطأ المعياري s_r يعطى بالقيمة $\sqrt{\frac{1 - r^2}{n - 2}}$.

(ii) تحديد مستوى المعنوية α.

(iii) باستخدام جدول t رقم (٢) آخر الكتاب لتحديد قيمة $t_{\upsilon,\alpha}$ الحرجة والتي تستخدم في تحديد منطقة قبول ورفض H_0 ونوضح طريقة الحساب من خلال المثال التالي.

مثال (١٣,٣)

باستخدام بيانات مثال (١٣,٢) لدرجات الطلاب اختبر معنوية معامل الارتباط r المحسوب من العينة عند مستوى معنوية $\alpha = 0.05$.

الحل

أولا: صياغة الفرض الإحصائي

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

ثانيا: اختيار الإحصاءة ثم حسابها

تحسب قيمة t_0 من العلاقة (١٣,٤) كما يلي:

تقرب قيمة r ولتكن $r = 0.92$ ودرجات الحرية هي $\upsilon = 8-2 = 6$ ثم نحسب t_0 كما يلي:

$$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1 - r^2}} = \frac{0.92\sqrt{8-2}}{\sqrt{1 - (0.92)^2}} = \frac{2.2535305}{\sqrt{0.1536}}$$

$$= \frac{2.2535305}{0.3919183} = 5.75$$

ثالثا: تحديد منطقة قبول ورفض H_0:

حيث تكون فترة قبول H_0 هي $(-t_{6,0.025}\ ,\ +t_{6,0.025})$ ومن جــدول t رقم (٢) آخر الكتـاب فإن فترة قـبـول H_0 هـي $(-2.447, 2.447)$ ونرفض H_0 خارج هذه الفترة.

رابعا: القرار الإحصائي

وحيث $t_0 = 5.75$ المحسوبة واقعة خارج فترة قبول H_0 والتي هي $(-2.447, 2.447)$ فإن القرار هو رفض H_0 أي أنّ المتغيران مرتبطان خطيا عند مستوى معنوية $\alpha = 0.05$.

(١٣,٥) معامل الارتباط للرتب لسبيرمان

Rank Coefficient of Correlation for Spearman

هناك معامل ارتباط آخر يستخدم بشكل كبير في التطبيقات الإحصائية التي تكون فيها البيانات مقاسة على شكل رتب، مثل تقديرات الطلاب في مادة معينة،

أو المستوى الدراسي لمجموعة من الطلاب والتي لا يمكن في هذه الحالة استخدام معامل ارتباط برسون فيها، والذي كما رأينا يحتاج إلى بيانات رقمية. معامل الارتباط هذا يعتمد على الرتب (حتى في حالة البيانات الرقمية يمكن ترتيبها واستخدام هذا المعامل فيها)، ويسمى معامل الارتباط للرتب لسبيرمان (Spearman) ويرمز له r_s ويقاس كما يلي:

نفرض أنه لدينا عينة حجمها n من مجتمع يراد دراسته ولتكن هذه العينة موضوعة في شكل أزواج مرتبة $(x_i, y_i), i = 1,2,3,\dots,n$ والآن نرتب القراءات (x_i, y_i) تصاعديا أو تنازليا مع بقاء كل قراءة في مكانها وإعطاء رتب $s_1, s_2, \dots, s_n$ للقراءات $x_1, x_2, \dots, x_n$ والرتب $t_1, t_2, \dots, t_n$ للقراءات $y_1, y_2, \dots, y_n$، والآن نشكل الفروق $d_1 = s_1 - t_1, d_2 = s_2 - t_2, \dots, d_n = s_n - t_n$ ومن هذه الفروق $d_i, i = 1,2,\dots,n$ نحسب معامل الارتباط للرتب r_s بالعلاقة (١٣,٥) كما يلي:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \qquad (١٣,٥)$$

ونوضح طريقة الحساب بالمثال التالي:

مثال (١٣,٤)

احسب معامل الارتباط للرتب لدرجات الطلاب في الإحصاء والرياضيات في مثال (١٣,٢) السابق.

الحل

لتسهيل الحل نكون جدول (١٣,٣) حيث نعطي قيم x و y رتب s و t تنازليا كما يلي:

جدول (٣, ١٣)

x	y	رتبة x_s	رتبة y_t	d = s-t	d^2
13	15	4	2.5	1.5	2.25
9	7	7	8	-1	1
19	17	1	1	0	0
15	15	3	2.5	0.5	0.25
11	10	5.5	5.5	0	0
8	9	8	7	1	1
16	14	2	4	-2	4
11	10	5.5	5.5	0	0
				Σd^2	8.5

ومن جدول (٣, ١٣) نحسب r_s من الصيغة (٥, ١٣) كما يلي:

$$r_s = 1 - \frac{6(8.5)}{8(64\text{-}1)} = 1 - 0.11904$$

$$= 0.8988 \cong 0.9$$

ملاحظة

في حالة القراءات المكررة يعطي كل منها متوسط الرتب للقراءات المكررة. فمثلا في مثال (٤, ١٣) السابق في حالة المتغير y كانت القراءات y ورتبتها t (ترتب تنازليا) كما في الجدول التالي:

القراءات y	15	7	17	15	10	9	14	10
الرتبة t	2	8	1	3	5	7	4	6

نلاحظ في الجدول السابق أن القراءة 15 مكررة مرتين برتب 2 و 3 والقراءة 10 مكررة مرتين برتب 5 و 6 ولهذا فإن القراءتان 15 تأخذ الرتبة $2.5 = \frac{3+2}{2}$ لكل منهما، والقرائتان 10 تأخذ الرتبة $5.5 = \frac{6+5}{2}$ لكل منهما فيكون الشكل النهائي

للرتب الموضحة في الجدول السابق بالجدول التالي:

القراءات y	15	7	17	15	10	9	14	10
الرتبة t	2.5	8	1	2.5	5.5	7	4	5.5

وهذه هي الرتب التي استخدمناها في جدول الحل (١٣,٣) السابق. لاحظ أن القراءة 14 والتي تلي 15 أخذت الرتبة 4 وليست الرتبة 3 وكذلك القراءة 9 والتي هي تلي القراءة 10 أخذت الرتبة 7 وليست الرتبة 6.

(١٣,٦) اختبار الفروض حول معامل الارتباط للرتب

Testing of Hypothesis for Rank Coefficient of Correlation

بعد حساب معامل الارتباط للرتب r_s من العينة المشاهدة مثال (١٣,٤) السابق يبقى السؤال حول معنوية هذه القيمة، وهل هناك ارتباط بين المتغيرين x و y في المجتمع الذي سحبت منه تلك العينة؟ وكما سبق لكي نجيب على هذا السؤال نجري اختبار الفروض كما يلي:

أولا: صياغة الفرض الإحصائي

$$H_0 : \rho = 0$$

$$H_1 : \rho > 0 \text{ أو } \rho < 0$$

نلاحظ الجدول (١٣,٤) التالي للقيم الحرجة للإحصاءة r_s بطرف واحد وعندما $H_1 : \rho > 0$ في حالة العلاقة الطردية الموجبة وعندما $H_0 : \rho < 0$ في حالة العلاقة العكسية السالبة.

ثانيا: اختيار الإحصاءة ثم حسابها:

الإحصاءة هي r_s التي تحسب من الصيغة (١٣,٥) السابقة.

ثالثا: تحديد فترة قبول ورفض H_0:

جدول (٤ ,١٣)

n	القيم الحرجـــة $r_{s,n,\alpha}$	
	$\alpha = 0.05$	$\alpha = 0.01$
4	1.000	--
5	0.900	1.000
6	0.829	0.943
7	0.714	0.893
8	0.643	0.833
9	0.600	0.783
10	0.564	0.746
12	0.506	0.712
14	0.465	0.645
16	0.325	0.601
18	0.399	0.564
20	0.377	0.534
22	0.359	0.508
24	0.343	0.485
26	0.329	0.465
28	0.317	0.448
30	0.306	0.432

حيث تحدد فترة قـبـول H_0 عند مستوى معنـويـة α بالفتـرة $(0, r_{s,\alpha,n})$ حيث $r_{s,\alpha,n}$ هي قيمة r_s الحرجة عندما يكون حجم العينـة n ومستـوى المعنوية α، جـدول (٤ ,١٣) القيم الحرجـة لمعامل ارتباط الرتب r_s عند مستوى معنوية $\alpha = 0.05$ و $\alpha = 0.01$ بذيل واحد.

مثال (١٣ , ٥)

أوجد معنوية ارتباط الرتب r_s المحسوب في مثال (١٣ , ٤) السابق الخاص بدرجات الطلاب في مادتي الرياضيات والإحصاء.

الحل

أولا: صياغة الفرض الإحصائي

$$H_0 : \rho = 0$$

$$H_1 : \rho > 0$$

ثانيا: اختيار الإحصاءة ثم حسابها

الإحصاءة هي r_s المحسوبة بالصيغة (١٣ , ٥) السابقة حيث كانت قيمتها في مثال (١٣ , ٤) السابق هي $r_s = 0.9$.

ثالثا: تحديد منطقة قبول ورفض H_0

تحدد فترة قبول H_0 من جدول القيم الحرجة لمعامل ارتباط الرتب $r_{s, \alpha, n}$ (١٣ , ٤) حيث فترة القبول H_0 هي $(0, r_{s, \alpha, n})$ أي الفترة $(0, r_{s, 0.05, 8})$ حيث من جدول رقم (١٣ , ٤) فإن هذه الفترة هي $(0, 0.643)$ ومنطقة رفض H_0 هي خارج تلك الفترة عند مستوى معنوية $\alpha = 0.05$.

رابعا: القرار الإحصائي

حيث قيمة $r_s = 0.9$ المحسوبة واقعة خارج منطقة قبول H_0 عند مستوى معنوية $\alpha = 0.05$ والتي هي $(0, 0.643)$ فإننا نرفض H_0 أي أنه يوجد ارتباط طردي بين درجات الامتحان للرياضيات والإحصاء في مثال (١٣ , ٤) السابق.

مثال (١٣,٦)

حصـل 6 طلاب على التقديرات التـالية في مادتي الإحصاء والرياضيـات (A,B), (C,B), (C,D), (C,C), (B,A), (D,E) هل هناك ارتباط بين تقديرات المادتين؟ ثم اختبر معنويته عند مستوى معنوية $\alpha = 0.01$.

الحل

نكون جدول (١٣,٥) للحل كما يلي:

جدول (١٣,٥)

x	y	رتبة x_s	رتبة y_t	$d = s-t$	d^2
A	B	1	2.5	−1.5	2.25
C	B	4	2.5	1.5	2.25
C	D	4	5	-1	1
C	C	4	4	0	0
B	A	2	1	1	1
D	E	6	6	0	0
					$\Sigma d^2=6.5$

ثم نحسب r_s كما يلي:

$$r_s = 1 - \frac{6 \Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6(6.5)}{6(36-1)} = 0.814$$

أولا: صياغة الفرض الإحصائي

$$H_0 : \rho = 0$$

$$H_1 : \rho > 0$$

ثانيا: اختيار الإحصاءة ثم حسابها

الإحصاءة هي r_s المحسوبة في المثال وهي $r_s = 0.814$.

ثالثا: تحديد منطقة قبول ورفض H_0 عند مستوى معنويــة $\alpha = 0.01$. حيث تكون فترة قبول H_0 هي $(0, r_{s, 0.01, 6})$ ومن جدول القيم الحرجة $r_{s, n, \alpha}$ جدول (٤ , ١٣) الأسبق فإن هذه الفترة هي (0, 0.943) .

رابعا: القرار الإحصائي:

حيث إن $r_s = 0.814$ المحسوبة واقعة داخل فترة قــبــول H_0 عند مستــوى معنـــوية $\alpha = 0.01$ حيث هذه الفترة هــي (0, 0.943) فإن القرار هو اننا لا نستطيع رفض H_0 لوقوع قيمة $r_s = 0.814$ داخل فترة قبــول H_0 أي أنه لا يوجد ارتباط بين تقدير الطلاب للمادتين عند مستوى معنوية $\alpha = 0.01$.

ملاحظة

عند تفسيرنا لمعامل الارتباط سواء كان لبيرسون أو لسبيرمان فإنــه يــجــب توخي الحذر في ذلك إذ أن وجود ارتباط بين المتغيرين لا يعني أن هناك علاقة سبب وتأثير (cause - and effect) بمعنى أن التغير في المتغير الأول لا يعني إطلاقا أنه سبب التغير في المتغير الثاني. فمثلا لو كانت x ترمز لمرتبات ضباط الشرطة في المملكة العربية السعودية خلال العشر سنوات الأخيرة، و y ترمز لاستهلاك المرطبات في المملكة في العشر سنوات الأخيرة فإن عينة من هذين المتغيرين سوف تبين أن هناك علاقة ارتباط قوية بينهما، ولكن منطقيا لا يوجد سبب أدى إلى زيادة استهــلاك المرطبات مع زيادة رواتب ضباط الشرطة. وأمثلة أخرى للسبب والتأثير هو عــدد الأميال التي تقطعها سيارة ما مع مقدار التآكل في الاطارات لهذه السيارة، وزمن التعرض للإشعاع والإصابة بمرض خطير، وطول فترات القيادة للسيارات في شوارع مدينة الرياض وعدد الحوادث التي يتعرض لها الشخص. . . الخ.

(١٣,٧) معامل الاقتران ومعامل التوافق

Coefficient of Contingency & Coefficient of Association

يستخدم معامل الاقتــران ومعامل التوافق لقياس قوة الارتباطات للبيانــات الوصفية بين الصفات المميزة ولكن لا نستطيع ترتيبها مثل الحالة الاجتماعية (أعزب - متزوج - أرمل - مطلق) وكذلك لون البشرة (أبيض - أسود -. . . . .) وكذلك الجنسية (سعودي - مصري. . . الخ). وفي مثل تلك الحالات يفشل قيــاس قــوة الارتباط باستخدام معامل الارتباط لبيرسون لاحتياجه لبيانات كمية كما يفشل أيضا مقياس سبيرمان لاحتياجه لبيانات وصفية لها صفة التراتيب، مثل تقديرات الطلاب مثلا، وفي حالة معامل الاقتران تقسم كل من الظاهرتين المطــلــوب دراســة قــوة الارتباط بينهما إلى صفتين فقط، على سبيل المثال التعليم (متعلم - غير متعلــم) والتدخين (مدخن - لا يدخن) وأما معامل التوافق فإنه يقيس قوة الارتبــاط بــين صفتين لكل منهما له أكثر من صفة مميزة لون الشعر (أسود - بني - أشقــر. . .) ولون البشرة (أبيض - أسود - أسمر. . .) ونوضح حساب كل من معاملي الاقتران والتوافق كما يلي:

أولا: معامل الاقتران

تقسم الصفة الأولى x والصفة الثانية y إلى الزوجين المرتبين (A,B), (C,D) كما هو موضح في جدول (١٣,٦) كما يلي:

جدول الاقتران (١٣,٦)

الصفة x / الصفة y	الصفة الأولى	الصفة الثانية
الصفة الأولى	A	B
الصفة الثانية	C	D

ويحسب معامل الاقتران r_c بالصيغة (٦ , ١٣) التالية:

$$r_c = \frac{AD - CB}{AD + CB} \qquad (١٣ , ٦)$$

مثال (٧ , ١٣)

عند دراسة علاقة التدخين بالتعليم في احدى المصانع أخذت عينة مكونة من 50 عاملا وكانت النتائج موضحة بالجدول التالي:

التدخين / التعليم	لا يدخن	يدخن
متعلـــــم	25	5
غير متعلم	10	10

احسب معامل الاقتران r_c بين التدخين والتعليم

الحل

باستخدام الصيغة (٦ , ١٣) فإن معامل الاقتران هو كما يلي:

$$r_c = \frac{(25(10) - 5(10)}{25(10) + 5(10)} = \frac{200}{300} = 0.67$$

وهو ارتباط متوسط.

ثانيا: معامل التوافق

يستخدم عندما تكون بيانات الظاهرتين التي لدينا عبارة عن بيانات وصفيــة لكل منهما أو وصفية لإحداهما وكمية للأخرى وكانت مقسمة إلى أكثر من نوعين (أي أن الجدول يحتوي على أكثر من أربع خانات) فإن معامل الاقتران السابــق لا

يصلح في هذه الحالة ويستخدم مقياسا آخر هو معامل التوافـق r_a (أوجده كرامير Cramer, 1946) ولحساب معامل التوافق نفرض أنه لدينا الظاهرة x لها r من الصفات والظاهرة الثانية y لها s من الصفات ويوضح جدول توافق الظاهرتين x و y بالجدول (١٣,٧)التالي:

جدول التوافق (١٣,٧)

الظاهرة y / الظاهرة x	y_1	y_2	... y_s	المجموع
x_1	f_{11}	f_{12}	f_{1s}	$f_{1.}$
x_2	f_{21}	f_{22}	f_{2s}	$f_{2.}$
x_r	f_{r1}	f_{r2}	f_{rs}	$f_{r.}$
المجموع	$f_{.1}$	$f_{.2}$	$f_{.s}$	$n=f_{..}$

ثم يحسب المقدار B من الجدول (١٣,٧) كما يلي:

$$B = \frac{(f_{11})^2}{f_{.1}(f_{1.})} + \frac{(f_{12})^2}{f_{.2}(f_{1.})} + \dots + \frac{(f_{rs})^2}{f_{.s}(f_{r.})} \qquad (١٣,٧)$$

وبالتعويض السابق ثم حساب معامل التـوافـق r_a بالصيغة (١٣,٨) التالية كمـا يلي:

$$r_a = \sqrt{\frac{B-1}{B}} \qquad (١٣,٨)$$

ونوضح ذلك بالمثال التالي.

مثال (٨ , ١٣)

عند دراسة العلاقة بين الرائحة ولون الزهور لعينة من 30 زهرة كانت لدينا النتائج التالية:

الرائحة y / اللون x	بدون رائحة	له رائحة	المجموع
أصفر	6	4	10
أبيض	7	2	9
أحمر	6	5	11
المجموع	19	11	30

احسب معامل التوافق r_a بين اللون والرائحة للزهور.

الحل

نحسب قيمة B من الجدول بالمثال باستخدام الصيغة (٧ , ١٣) كما يلي:

$$B = \frac{6^2}{19(10)} + \frac{7^2}{19(9)} + \frac{6^2}{19(11)} + \frac{4^2}{11(10)} + \frac{2^2}{11(9)} + \frac{5^2}{11(11)} = 1.05$$

ويحسب معامل التوافق r_a بالصيغة (٨ , ١٣) السابقة كما يلي:

$$r_a = \sqrt{\frac{B - 1}{B}} = \sqrt{\frac{1.05 - 1}{1.05}} = 0.22$$

يلاحظ أن معامل التوافق $r_a = 0.22$ يبين أن قوة الارتباط بين اللون والرائحة للزهور تكون ضعيفة في هذا المثال.

ملاحظة

يمكن الحصول على معامل التوافق r_a المعطى بالصيغة (١٣,٨) بصيغة أخرى تستخدم مربع كاي(χ_υ^2) وتعطى بالعلاقة (١٣,٩) التالية:

$$r_a = \sqrt{\frac{\chi_\upsilon^2}{\chi_\upsilon^2 + n}} \qquad (١٣,٩)$$

حيث χ_υ^2 تحسب بالصيغة (٨,٤) من الفصل الثامن السابق. وباستخدام القيمة n وتمثل مجموع التكرارات في خلايا جدول التوافق وهي عبارة عن حجم العينة. ولقد أثبت كرامير أن معامل التوافق بالصيغة (١٣,٨) يعطي نفس القيمة لمعامل التوافق المعطى بالصيغة (١٣,٩) ويمكن التحقق من ذلك بحساب معامل التوافق بالصيغة (١٣,٩) في مثال (١٣,٨) ويترك ذلك كتمرين للطالب أو الباحث.

(١٣,٨) الانحدار الخطي البسيط
Simple Linear Regression

في البنود السابقة من هذا الفصل تعرضنا لأمثلة لمتغيرات يؤدي التغير في أحدها إلى التغير في الآخر، وقلنا إن سبب التغير في المتغير الأول أدى إلى تغير أو أثر في المتغير الثاني، ومن ثم وضعنا هذا على شكل مقياس للارتباط واستعرضنا مقاييس الارتباط ثم اختبار الفروض لبعض هذه المقاييس. يبقى الآن السؤال: كيف نستفيد من معرفة إذا ما كانت هناك علاقة بين متغيرين أو ارتباط بينهما؟ وكيف نستطيع استغلال هذا الارتباط للتنبؤ عن حالة أحدهما بمعرفة حالة الآخر؟ هنا يأتي دور الانحدار(regression) أوخط الانحـدار(regression line) والذي يعطي علاقة خطية رياضية تربط بين أحد المتغيرات والذي نسميه المتغير المستقل مع المتغير الآخر الذي نسميه المتغير التابع. بمعنى آخر بينهما ارتباط يعطي مقياس لقوة العلاقة الخطية بين المتغيرين. فإن الانحدار يعطي طريقة عددية لقياس هذه العلاقة الخطية

بين متغيرين. فمثلا في مثالنا عن درجات الطلاب في الإحصاء والرياضيات وجدنا أن هناك علاقة خطية قوية بين درجات الامتحان في المادتين. الآن بإمكاننا استخدام درجات الامتحان في مادة ما للتنبؤ عن درجة الامتحان التي يمكن الحصول عليها في المادة الأخرى، وهنا كل من المتغيرين يصلح لأن يكون متغيرا مستقلا أو تابعا. كمثال آخر معروف أن ضغط الدم السفلي يدل دلالة قوية أو يرتبط ارتباطا قويا بحالة المريض الصحية (كمقياس وحيد وسريع)، وهنا يكون ضغط الدم السفلي هو المتغير المستقل وحالة المريض الصحية هي المتغير التابع.

إذن هنا سوف نستعرض تحليل الانحدار والذي يقوم باشتقاق معادلة خطية (معادلة خط مستقيم) والتي منها نستطيع ايجاد قيمة y (المتغير التابع) إذا أعطينا قيمة x (المتغير المستقل) أي التنبؤ عن y بمعرفة x. هنا نود أن نقول إن قيمة y المتنبأ بها هي قيمة وسطية وليست دقيقة أو صحيحة تماما، بمعنى أن القيمة x المعطاة تكون قيمة y الناتجة من المعادلة هي قيمة في المتوسط وليست قيمة دقيقة مئة في المئة.

لايجاد معادلة خط الانحدار البسيط لعينة معطاة من مشاهدات على شكل أزواج مرتبة لمتغيرين x و y على الشكل التالي:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

ولنفترض أننا وجدنا أنه هناك ارتباط معنوي بين المتغيرين. والآن نود ايجاد علاقة خطية تربط بينهما (بافتراض أن x المتغير المستقل و y المتغير التابع) وتصاغ هذه العلاقة كما يلي:

$$y = a + b\,x \qquad (١٣,١٠)$$

وهنا يتم اختيار المعاملات a و b بحيث إن هذا الخط يعطي أفضل تطبيق للمشاهدات كما سنبين في الفقرة التالية.

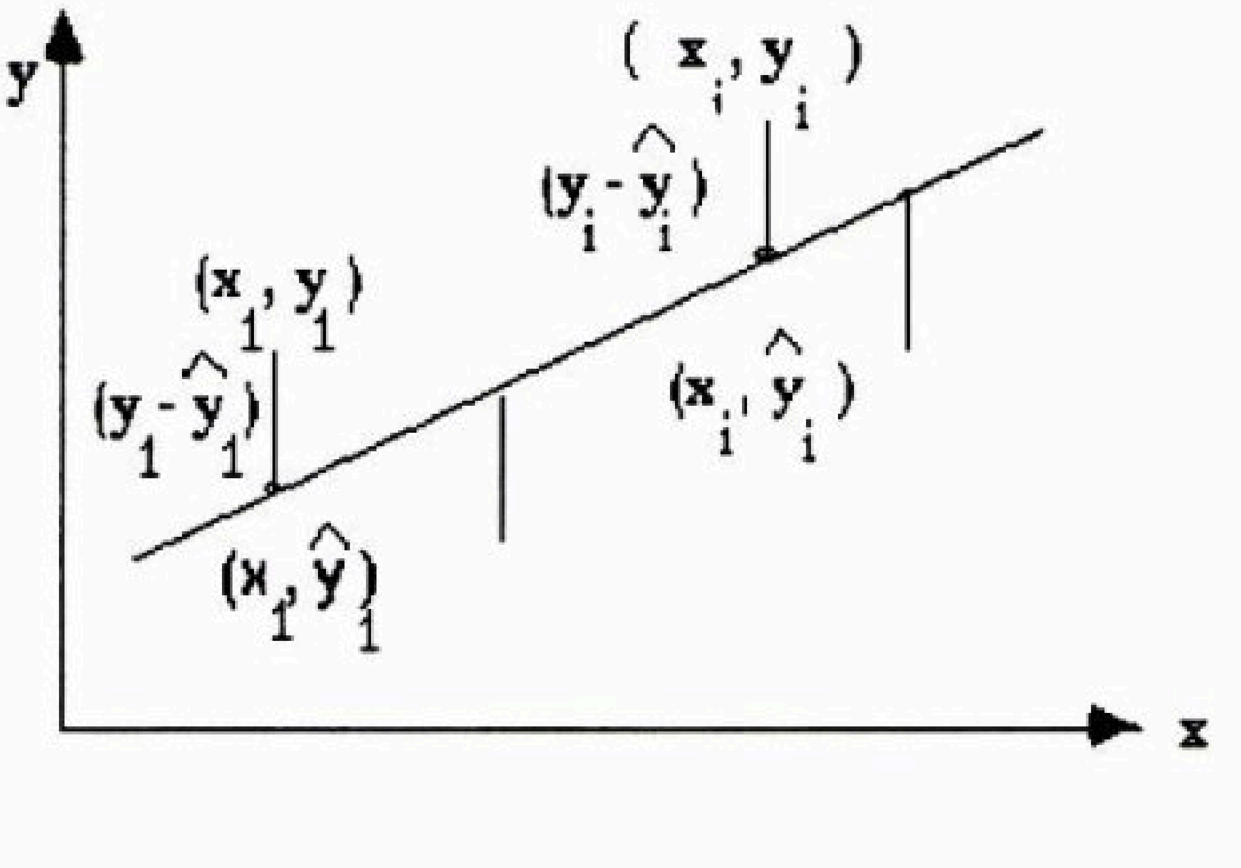
(١٣,٨,١) طريقة المربعات الصغرى

إذا أعطينا مجموعة من المشاهدات على شكل أزواج مرتبة (x, y) فإنه

بإمكاننا أن نجد خط مستقيم يعطي أفضل تطابق للمشاهدات المعطاه. فمثلا إذا وجدنا أنه توجد علاقة أو ارتباط خطي مناسب فإننا في هذه الحالة سوف نفترض أن خط مستقيم يمكن أن يمثل هذه المشاهدات بشكل مناسب، وهنا سوف نبحث عن أفضل خط مستقيم وذلك باستخدام طريقة المربعات الصغرى (least squares method) لنفترض أن

$$\hat{y} = a + b\,x \qquad (١٣,١١)$$

هي معادلة خط مستقيم حيث $\hat{y}$ (تقرأ y هات) تمثل القيم المتنبىء عنها للمتغير y والتي تنتج من قيمة معينة x. طريقة المربعات الصغرى تقوم على إيجاد a و b بحيث إن المجموع $\Sigma(y - \hat{y})^2$ يكون أقل ما يمكن أنظر الشكل (١٣,٩) التالي:



شكل (١٣,٩)

نلاحظ في شكل (١٣,٩) أن المقدار $(y_i - \hat{y}_i)$ يمثل مقدار الانحراف للمشاهدة (x_i, y_i) عن الخط المستقيم، وطبيعي نود أن تكون هذه الانحرافات أقل ما يمكن لكي تعطي أفضل خط مستقيم يمر بشكل مناسب بالقرب من جميع القراءات. لاحظ أن الانحرافات $(y_i - \hat{y}_i), i = 1,2,\dots,n$ تكون بعضها موجبة والأخرى سالبة حسب موقعها فوق أو تحت الخط المستقيم، لهذا فإن الانحرافات

السالبة تلغي الانحرافات الموجبة، ولهذا أخذنا مربع الانحرافات حتى نتجنب هذه المشكلة، وهنا يهمنا أن يكون مجموع مربع جميع الانحرافات أقل ما يمكن، وعلى هذا الأساس يكون اختيار a و b. بالتعويض عن قيمة $\hat{y}$ من الصيغة (١١ ,١٣) فإن معادلة مجموع المربعات تصبح على الشكل التالي:

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a - b x_i)^2 \tag{١٣, ١٢}$$

نريد أن نصغر العلاقة (١٢ ,١٣) السابقة بالنسبة للمعاملات a و b، وهنا يسعفنا حساب التفاضل والذي يقول لتصغير علاقة بالنسبة لمتغير أو متغيرين فإننا نشتق هذه العلاقة بالنسبة للمتغير اشتقاق كامل، أو اشتقاق جزئي في حالة المتغيرين وتساوي المعادلة أو المعادلات الناتجة بالصفر، فنحصل على ما يسمى بالمعادلات الطبيعية، والتي بحلها بالنسبة للمتغيرات تعطي النتيجة المطلوبة (تصغيرا أو تكبيرا). الآن نشتق العلاقة جزئيا بالنسبة للمقدارين a و b فنجد أن:

$$\frac{\partial}{\partial a} \sum_{i=1}^{n} (y_i - a - b x_i)^2 = -2 \sum_{i=1}^{n} (y_i - a - b x_i)$$

$$\frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - a - bx_i)^2 = -2 \sum (y_i - a - bx_i) x_i$$

وبوضع المعادلات السابقة مساوية للصفر ومع قليل من الاختصارات نحصل على المعادلتين التاليتين وهما:

$$n a + b (\Sigma x_i) = \Sigma y_i \tag{١٣, ١٣}$$

$$n(\Sigma x_i) + b (\Sigma x_i^2) = \Sigma x_i y_i \tag{١٣, ١٤}$$

وبحل المعادلتين (١٣ , ١٣) و (١٤ , ١٣) السابقتين نحصل على قيمتي a و b كما يلي:

$$b = \frac{n \Sigma x y - (\Sigma x)(\Sigma y)}{n \Sigma x^2 - (\Sigma x^2)} \qquad (١٣ , ١٥)$$

$$a = \overline{y} - b \overline{x} \qquad (١٣ , ١٦)$$

حيث
$$\overline{y} = \frac{\Sigma y}{n} , \quad \overline{x} = \frac{\Sigma x}{n}$$

والعلاقتين (١٥ , ١٣) و (١٦ , ١٣) تعطينا قيمتا b و a واللتان تمثلان أفضل تطابق للمشاهدات مع الخط المستقيم $y = a + b x$. نلاحظ أنه يتم حساب القيمة b من الصيغة (١٥ , ١٣) ثم نستخدمها في حساب a في الصيغة (١٦ , ١٣).

مثال (٩ , ١٣)

أوجد معادلة خط انحدار درجات امتحان الطلاب في مادة الرياضيات (y) إذا علم أن درجاتهم في مادة الإحصاء (x) في مثال (٢ , ١٣) السابق. ثم أوجد درجة أحد الطلاب المتوقعة في الرياضيات إذا حصل على درجة في الإحصاء مقدارها 13 درجة.

الحل

من جدول (٢ , ١٣) الخاص بالمثال (٢ , ١٣) لدرجات الطلاب في الإحصاء والرياضيات كانت لدينا المجاميع التالية:

$$n = 8, \ \Sigma x = 102, \ \Sigma y = 97, \ \Sigma xy = 1322, \ \Sigma x^2 = 1398$$

ويمكن حساب b من الصيغة (١٥ , ١٣) كما يلي:

$$b = \frac{n\Sigma xy - \Sigma x\Sigma y}{n\Sigma x^{2} - (\Sigma x)^{2}} = \frac{8(1322) - (102)(97)}{8(1398) - (102)^{2}}$$

$$= \frac{682}{780} = 0.8744$$

ثم نحسب a من الصيغة (١٣,١٦) كما يلي:

$$a = \overline{y} - b\,\overline{x}$$

$$= \frac{97}{8} - \frac{102}{8}(0.8744)$$

$$= 12.125 - (12.75)(0.8744) = 0.9764$$

وعليه فإن معادلة خط الانحدار المطلوبة هي:

$$y = 0.9764 + 0.8744\,x$$

ونعوض عن قيمة $x = 13$ في المعادلة السابقة نحصل على القيمة المتوقعة لدرجة الطالب في مادة الرياضيات y كما يلي:

$$y = 0.9764 + 0.8744(13)$$

$$= 12.344 \quad \text{درجة}$$

أي نتوقع (في المتوسط) أن يحصل هذا الطالب على 12 درجة تقريبا في مادة الرياضيات.

مثال (١٣,١٠)

في عينة عشوائية مكونة من 8 طلاب في الثانوية العامة لهذا العام وجدت أطوالهم بالسنتيميتر وأوزانهم بالكيلوجرام على النحو التالي (في الكشف الطبي للالتحاق بالجامعة):

الطول x	165	165	157	170	175	165	155	170
الوزن y	48	57	50	54	64	61	43	59

أوجد معادلة للتنبؤ عن وزن طالب ثانوية عامة لهذا العام معتمدا على طوله.

الحل

قبل أن نوجد معادلة التنبؤ يجب أن نقرر فيما إذا كانت الأطوال والأوزان السابقة مرتبطة ولهذا سوف نحسب معامل الارتباط r ونكون لذلك جدولا للحل (٨,١٣) كما يلي:

جدول (٨,١٣)

x	y	xy	x^2	y^2	
165	38	7920	27225	2304	
156	57	9405	27225	3249	
157	50	7850	24649	2500	
170	54	9180	28900	2916	
175	64	11200	30625	4096	
165	61	10065	27225	3712	
155	43	6665	24025	1849	
170	59	10030	28900	3481	
1322	436	72315	218774	24116	المجموع

نحسب معامل ارتباط بيرسون r والذي بالصيغة (٢,١٣) السابقة وهي:

$$r = \frac{SS(xy)}{\sqrt{SS(x).SS(y)}}$$

باستخدام المجاميع التي بالجدول (٨,١٣) نحسب الكميات التالية:

$$SS(xy) = \Sigma\, xy - \frac{\Sigma x\, \Sigma y}{n} = 72315 - \frac{(1322)(436)}{8} = 266$$

$$SS(x) = \Sigma\, x^2 - \frac{(\Sigma x)^2}{n} = 218774 - \frac{(1322)(1322)}{8} = 313.$$

$$SS(y) = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 24116 - \frac{(436)(436)}{8} = 354$$

ثم نحسب معامل الارتباط r باستخدام االكميات السابقة كما يلي:

$$r = \frac{266}{\sqrt{(3136)(354)}} = \frac{266}{333.13509} = 0.7985$$

والآن نختبر معنوية الارتباط r المحسوب من العينة كما يلي:

أولا: صياغة الفرض الإحصائي:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

ثانيا: اختيار الإحصاءة ثم حسابها

نختار الإحصاءة $t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ ثم نحسبهـا كما يلي

حيث إن $n = 8$, $r = 0.7985$, $\upsilon = n-2 = 8-2 = 6$ فإن t_0 تحسب كما يلي:

$$t_0 = \frac{0.7985\sqrt{8-2}}{\sqrt{1-(0.7985)^2}} = \frac{1.9559}{0.6376} = 3.0676$$

ثالثا: تحديد منطقة قبول H_0 عند مستوى معنوية $\alpha = 0.05$

من طرفين هي($-t_{6,0.025}$, $t_{6,0.025}$) من جدول t (٢) آخر الكتاب فإن هذه الفترة تصبح (-2.447, +2.447) ومنطقة رفض H_0 خارج هذه الفترة.

رابعا: القرار الإحصائي

حيث إن t_0=3.0676 المحسوبة تقع خارج فترة قبــول H_0 والتي هـــــي (-2.447, 2.447) فإننا نرفــض H_0 أي أن معامل الارتباط بين وزن وطول الـطـلاب معنوي ويختلف من الصفر أي يوجد ارتباط بين x و y.

والآن نوجد خط انحدار الــوزن y على الطول x وذلك بإيجاد معامــلات

الخط المستقيم a و b. حيث يمكن وضع الصيغة (١٥ ,١٣) السابقة الخاصة بحساب معامل الانحدار على الصيغة (١٧ ,١٣) التالية:

$$b = \frac{SS(xy)}{SS(x)} \qquad (١٣, ١٧)$$

أي أن

$$b = 266/313.6 = 0.848$$

$$a = \overline{y} - b\,\overline{x} = \frac{436}{8} - (0.848)\ \frac{1322}{8} = -85.63$$

فإن معادلة خط انحدار الوزن (y) على الطول (x) تكون كما يلي:

$$y = -85.6 + 0.85\,x$$

قد نتساءل ما هو وزن طالب كان طوله 160 سم والإجابة هي:

$$y = -85.6 + 0.85(160) = 50.4 \text{ كجم}$$

أي أننا نتوقع وزن الطالب (في المتوسط) يكون 50 كجم تقريبا.

ملاحظات

● يمثل الميـل b التغير المتـوقع في y لتغير x وحدة فمثـلا في مثـال (١٠ ,١٣) السابق كانت b = 0.85 فلو أن طالبا زاد طوله 10 سم فإننا نتوقع زيادة 8.5 كجم على وزنه.

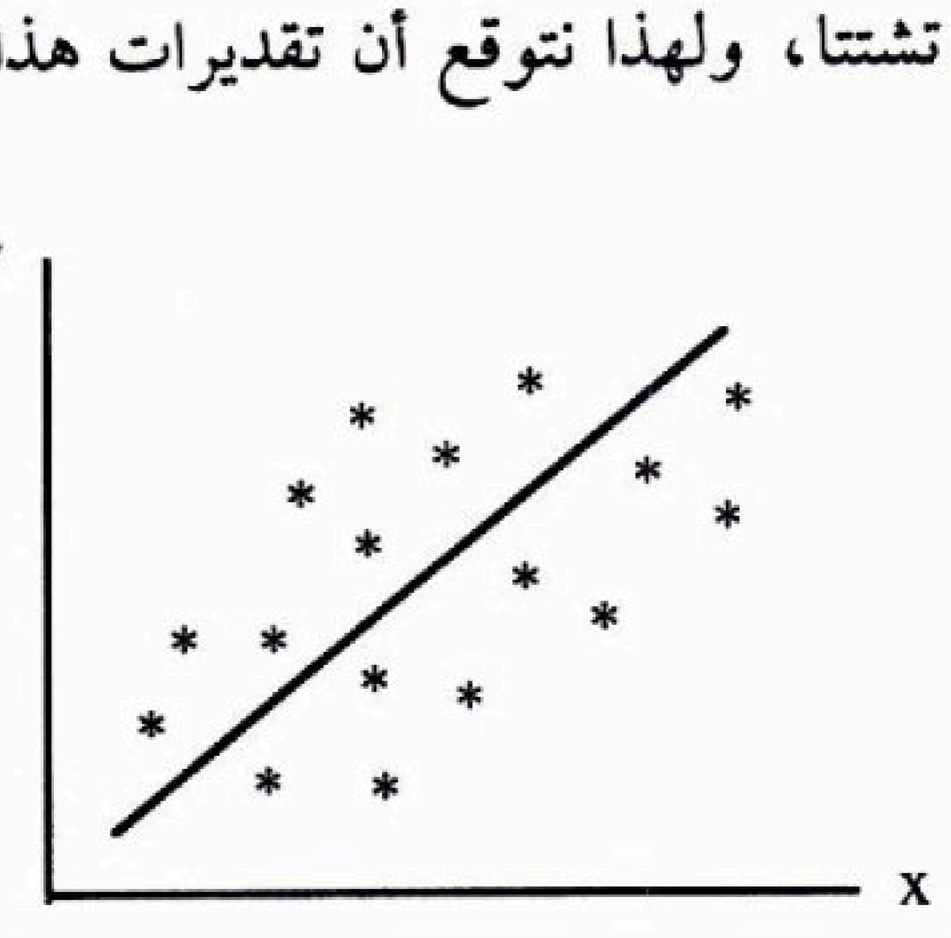
● يمثل a الموضع الذي يقطع فيه الخط المستقيم محور الصادات والذي لديه x = 0. هنا في مثالنا السابق عندما x = 0 (طالب لا طول له) يكون وزنه y = -86 أي وزنه بالسالب وهذا غير منطقي وغير صحيح، لهذا فإنه يجب توخي الحذر فـي استخدام خط الانحدار للتنبؤ، واستخدام قيم للمتغير x التي تكون ضمن مجال المشاهدات المعطاة في المجتمع ولا نحاول الخروج عن هذا المجال من أي جهـة،

فمثلا في مثال (١٠, ١٣) السابق مجال المتغير x (الطول) كان من 155 سم إلى 175 سم ولهذا يحتاط في عدم استخدام قيم أقل من 155 أو أكبر من 175 للتنبؤ.

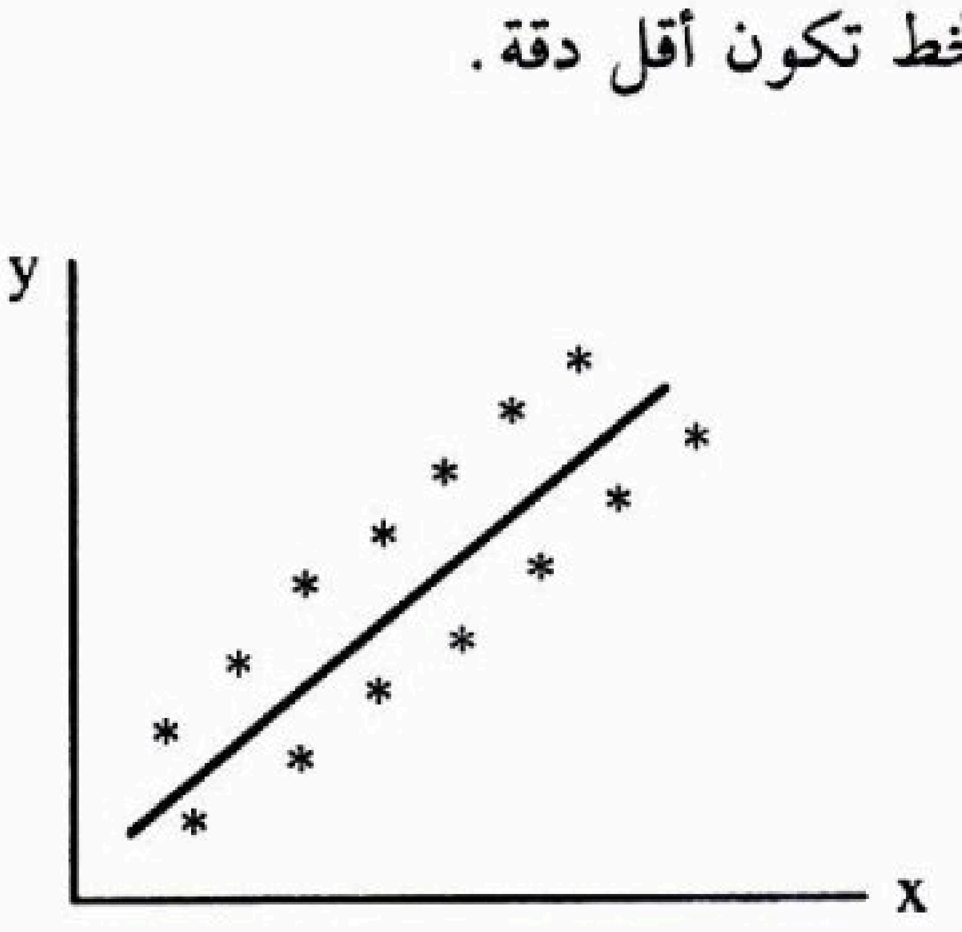
(١٣, ٨, ٢) الخطأ المعياري للتقدير

بعد أن وجدنا علاقة خطية رياضية في شكل خط انحدار بين المتغير المستقل x والمتغير التابع y يهمنا هنا قياس مدى اعتمادنا على المعادلة الرياضية التي وجدناها، فمثلا بالنظر إلى الشكلين (١٠, ١٣) و (١١, ١٣) التاليين:

نلاحظ أن خط الانحدار في شكل (١٠, ١٣) يعطي تقديرا أكثر دقة، وذلك لكون البيانات متقاربة أكثر من الخط أي أنها تنتشر بشكل أقل حول خط الانحدار بينما، في شكل (١١, ١٣) نلاحظ أن انتشار البيانات حول خط الانحدار أكـثـر تشتتا، ولهذا نتوقع أن تقديرات هذا الخط تكون أقل دقة.



شكل (١١, ١٣)



شكل (١٠, ١٣)

هنا سوف نعرف كمية عددية نسميها الخطأ المعياري للتقدير (Standard error of estimate)، ونرمز لها بالرمز s_e وهي تقيس التغير (أو الانتشار) للمشاهدات المعطاة حول خط الانحدار، وتعطى بالعلاقة التالية:

$$s_e = \sqrt{\frac{\Sigma(y - \hat{y})^2}{n - 2}} \qquad (١٣,١٨)$$

حيث y هي القيم المشاهدة المعطاة فـي الأزواج المـرتـبـة (x, y) و $\hat{y}$ هي القيمـة المحسوبة من المعادلة $\hat{y} = a + b x$.

مثال (١٣,١١)

من مثال (١٣,٢) لدرجات الطلاب في الإحصـاء (x) والرياضيـات (y). أوجد الخطأ المعياري للتقدير s_e لدرجات الطلاب في الإحصاء والرياضيات.

الحل

سبق وأن وجدنا معادلة انحدار درجات الرياضيات y على درجات الإحصاء x كما يلي:

$$y = 0.98 + 0.87\,x$$

نكون الآن جدول (١٣,٩) للحل كما يلي:

جدول (١٣,٩)

x	y	y = 0.98 + 0.87 x	$(y - \hat{y})$	$(y - \hat{y})^2$
13	15	12.34 = 0.98 + 0.87(13)	2.66	7.06
9	7	8.84 = 0.98 + 0.87(9)	-1.85	3.41
19	17	17.59 = 0.98 + 0.87(19)	-0.59	0.35
15	15	14.09 = 0.98 + 0.87(15)	0.91	0.82
11	10	10.59 = 0.98 + 0.87(11)	-0.59	0.35
8	9	7.97 = 0.98 + 0.87(8)	1.03	1.06
16	14	14.97 = 0.98 + 0.87(16)	-0.97	0.93
11	10	10.59 = 0.98 + 0.87(11)	-0.59	0.35
			0.01 = 0	14.33

نحسب الخطأ المعياري للتقدير s_e باستخدام العلاقة (١٣,١٨) السابقة والمقدار $\Sigma(y - \hat{y})^2$ من جدول (١٣,٩) كما يلي:

$$s_e = \sqrt{\frac{\Sigma(y - \hat{y})^2}{n - 2}} = \sqrt{\frac{14.33}{8-2}} = \sqrt{2.388333} = 1.545$$

ملاحظة

الكمية $\Sigma(y - \hat{y})^2$ تأخذ وقتا طويلا لحسابها فإذا افترضنا أننا لا نحتاج حساب الكميات $\hat{y}$ فإنه يوجد صيغة ابسط لحساب $\Sigma(y - \hat{y})^2$ بدون الحاجة لحساب $\hat{y}$ وهي:

$$\Sigma(y - \hat{y})^2 = \Sigma(y - a - b\,x)^2$$

وبقليل من الاختصارات فإنها تصبح كما يلي:

$$\Sigma(y - \hat{y})^2 = \Sigma\, y^2 - a\,\Sigma\, y - b\,\Sigma\, x\, y \qquad (١٣,١٩)$$

وبالتعويض بالعلاقة (١٣,١٩) في العلاقة (١٣,١٨) السابقة نحصل على الخطأ المعياري S_e بالعلاقة التالية

$$S_e = \sqrt{\frac{\Sigma y^2 - a\Sigma y - b\Sigma xy}{n - 2}} \qquad (١٣,٢٠)$$

والعلاقة (١٣,٢٠) سهلة الحساب عن العلاقة (١٣,١٨) التي عادة تستخدم.

وتحتاج الصيغتين (١٣,١٩) و (١٣,٢٠) فقط إلى حساب الكميات $\Sigma\, x\, y$, $\Sigma\, y$, $\Sigma\, y^2$ والتي تم حسابها عند ايجاد خط الانحدار كما نحتاج حساب قيم a و b وهي أيضا تم حسابها عند ايجاد خط الانحدار أيضا.

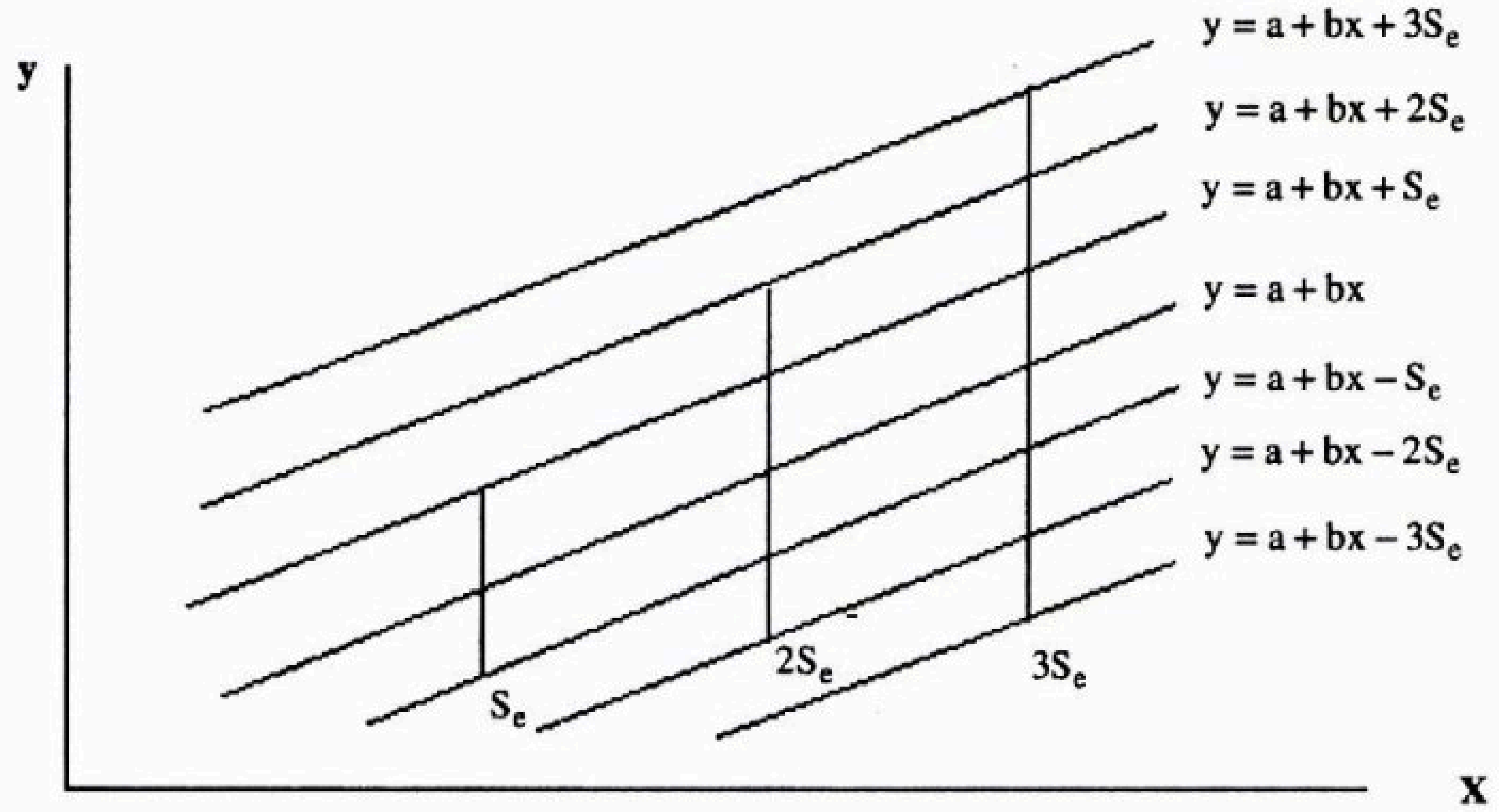
ولحساب $\Sigma(y - \hat{y})^2$ من العلاقة (١٣,١٩) يتم ذلك كما يلي:

$$\Sigma(y - \hat{y})^2 = 1265 - (0.977)(97) - (0.8744)(1322) = 14.2742$$

(الفرق بين هذه القيمة والقيمة السابقة الناتجة من العلاقة (١٨ , ١٣) راجع لأخطاء التقريب).

وهكذا وجدنا أن $s_e = 1.55$ وتفسير الخطأ المعياري للتقدير هو تماما كتفسير الانحراف المعياري فكلما كان الخطأ المعياري للتقدير كبيرا كلما كان تشتت المشاهدات حول خط الانحدار كبيرا. وهكذا لو افترضنا أن $s_e = 0$ فهذا يعني أن جميع المشاهدات تقع بشكل تام ودقيق على الخط، ولا توجد أي مشاهدة بعيدة عن الخط، وهكذا فإن معادلة التنبؤ سوف تكون مثالية في تقدير المتغير التابع y.

بافتراض أن النقاط المشاهدة موزعة توزيعا طبيعيا حول خط الانحدار فإننا نتوقع أن نجد 68٪ من النقاط في حدود $\pm 1\ s_e$ حول خط الانحدار وأن 95.5٪ من النقاط في حدود $\pm 2\ s_e$ حول خط الانحدار وأن 99.7٪ من النقاط في حدود $\pm 3\ s_e$ حول خط الانحدار ونوضح ذلك في شكل (١٢ , ١٣) كما يلي:



شكل (١٢ , ١٣)

لاحظ أن القياس يكون موازيا لمحور وليس عموديا على خط الانحـدار. سـوف نستخدم s_e الخطأ المعياري للتقدير لإيجاد فترات تنبؤ حول القيمة المقدرة $\hat{y}$ التي ستقع ضمنها القيمة الحقيقية y، فعلى افتراض أن القيم المشاهدة y موزعة توزيعا طبيعيا حول القيمة المقدرة $\hat{y}$ ، وكذلك إذا افترضنا أن تباين التوزيعات حول كل قيمة ممكنة $\hat{y}$ ثابت لا يتغير، فإننا [انظر شكل (١٢ , ١٣)] نكون متأكدين 95.5٪ بأن القيمة الحقيقيـة y سوف تقع ضمن خطأين معياريين من القيمة المقـدرة $\hat{y}$ ، وهذه الفترات تسمى بفترات التنبؤ التقريبية (لها نفس الدور الذي تلعبه فـتـرات الثقة التي درسناها سابقا).

مثال (١٢ , ١٣)

في مثال (١١ , ١٣) لدرجات الطلاب في الإحصـاء (x) والرياضيات (y). أوجد تقدير فترة تنبؤ 95.6٪ لدرجة طالب في الرياضيات إذا علم أنه حصل على 13 درجة في الإحصاء.

الحل

سبق وأوجدنا في مثال (٩ , ١٣) معادلة انحدار درجات الرياضيات (y) على درجات الإحصاء (x) وكانت المعادلة هي:

$$y = 0.98 + 0.87\,x$$

وسبق أن وجدنا الخطأ المعياري للتـنـبـؤ s_e في مثـال (١١ , ١٣) السابق وكانت $s_e = 1.55$.

والآن نحسب القيمة المتوقعة $\hat{y}$ للطالب في الرياضيات وهي:

$$\hat{y} = 0.98 + 0.87\,(13) = 12.3 \text{ درجة}$$

إذن يكون هناك تقدير فترة تنبؤ تقريبيـة 95.5٪ لدرجة الطالب في الرياضيـات وهي:

$$\hat{y} - 2 s_e \le y \le \hat{y} + 2 s_e$$

أي أن

$$12.3 - 2(1.55) \le y \le 12.3 + 2(1.55)$$

أي أن تقدير الفترة للتنبؤ لدرجة الإحصاء هي:

$$9.2 \le y \le 15.5$$

وهذا يعني أن الدرجة الحقيقية ستكون داخل هذه الفترة (9.2, 15.2) وذلك باحتمال 0.955.

ملاحظة

● النتيجة السابقة غير دقيقة تماما وذلك نظرا لصغر العينة $(n = 8)$ واستخدام التوزيع الطبيعي كتقريب يحتاج إلى حجم العينة $n \ge 30$. لتصحيح هذا يجب استخدام توزيع t بدلا من التوزيع الطبيعي ولكن هنا استخدمنا التوزيع الطبيعي لتسهيل الشرح.

● كما أسلفنا سابقا فترة التنبؤ السابقة تقريبية وذلك لاستخدامنا s_e بدلا من s_b الخطأ المعياري للتنبؤ حيث s_b تعطى بالعلاقة التالية:

$$s_b = s_e \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\Sigma x^2 - n \bar{x}^2}} \qquad (١٣,٢١)$$

أو بصورة تقريبية فإن S_b تعطي بالعلاقة التالية:

$$S_b = S_e / \sqrt{\sum (x - \bar{x})^2}$$

حيث x_0 المعطاة في الصيغة (١٣,٢١) هي القيمة المعطاة للمتغير المستقل التي يراد التنبؤ بها عن المتغير التابع y. (في مثال (١٣,١٢) السابق كانت $x_0 = 13$).

(١٣,٩) اختبار الفروض حول معامل الانحدار وتقدير فترة الثقة له

Testing of Hypothesis for Regression Coefficeint & Confidence Interval

لقد استخدمنا حتى الآن تحليل الانحدار لكي نوجد علاقة بين متغيرين وذلك على أساس معلومات من عينة ولكن البيانات من عينة تمثل فقط جزء من المجتمع الكلي، وبسبب هذا فإننا سوف نعتبر خط الانحدار الذي حصلنا عليه عبارة عن تقدير لخط انحدار المجتمع الذي هو طبعا غير معروف والذي نفترضه على الصورة التالية:

$$y = \gamma + \beta x \qquad (١٣,٢٢)$$

سوف نعطي الآن بعض التفسير لخط الانحدار (١٣,٢٢) على أساس مثال درجات الطلاب في الإحصاء والرياضيات. حيث إن درجة الطالب في الإحصاء تنتج من جزئين.

(i) **الجزء الأول** ناتج من معرفة غير رياضية للطالب (مثل التعاريف البسيطة) وهذا الجزء يتمثل بالحد γ وهو ما يحصل عليه الطالب حتى لو كانت معلوماته الرياضية لا شيء (لا يملك مهارة رياضية).

(ii) **الجزء الثاني** ناتج عن مهارة الطالب الرياضية أي معرفته بالرياضيات (نظريات، طرق عد، حساب تكامل وتفاضل.. الخ) وهذا الجزء يزداد ويتناسب مع زيادة مهارة الطالب الرياضية ويتمثل هذا بالحد βx. طبعا مهارات الطلاب ليست متساوية، كما أن السرعة البديهية وسرعة استخدام الطرق الرياضية ليست متساوية في الطلاب ذوي المهارات المتكافئة، لهذا فإن قيم البيانات المشاهدة لن تقع تماما على خط الانحدار المجتمع محل الدراسة بل بعضها سيقع فوق الخط وبعضها يقع تحت الخط، ولهذا فبدلا من أن البيانات تحقق المعادلة (١٣,٢٢) فإنها في الحقيقة تحقق المعادلة (١٣,٢٣) التالية:

$$y = \gamma + \beta x + e \qquad (١٣,٢٣)$$

حيث e هي انحرافات عشوائية عن خط انحدار المجتمع وهذه الانحرافات تساوي صفرا في المتوسط لأن بعضها يكون موجبا وبعضها سالبا (فوق خط

الانحدار وتحت خط الانحدار). سوف نرمز لتباين هذه الانحرافات أو الأخطاء العشوائية بالرمز s_e^2، والخطأ المعياري للتقدير s_e ما هو إلا تقديرا للمقدار σ_e الانحراف المعياري للأخطاء العشوائية.

بما أن خط انحدار العينة $y = a + bx$ هو عبارة عن تقدير لخط انحدار المجتمع محل الدراسة $y = \gamma + \beta x$ فإننا سوف نستخدمه للاستدلال (inference) عن خط انحدار المجتمع (١٣,٢٢)، أي أننا سوف نقوم ببعض الاستدلالات عن β لخط الانحدار الحقيقي معتمدين على معامل الانحدار b لخط الانحدار للعينة. كذلك المقدار γ باستخدام a. أي أننا سنقوم باختبار الفروض حول β و γ باستخدام b و a.

أولا: اختبار الفروض حول معامل الانحدار β.

يصاغ الفرض الإحصائي كما يلي:

$$H_0 : \beta_0 = 0$$

$$H_1 : \beta \neq 0$$

ثم تحسب الإحصاءة t_0 حيث υ تمثل درجات الحرية وهي $\upsilon = n - 2$ والإحصاءة t_0 تحت صحة H_0 تعطى بالصيغة (١٣,٢٤) التالية:

$$t_0 = \frac{b - \beta_0}{s_b} \qquad (١٣,٢٤)$$

حيث s_b تمثل الخطأ المعياري لمعامل الانحدار b ويعطى بالعلاقة (١٣,٢٥) التالية:

$$s_b = \frac{s_e}{\sqrt{\Sigma x^2 - n\bar{x}^2}} \qquad (١٣,٢٥)$$

وتحدد منطقة قبول H_0 بالفترة $(-t_{\upsilon,\alpha/2}, t_{\upsilon,\alpha/2})$ حيث الاختبار البديل H_1 من طرفين ومنطقة رفض H_0 خارج تلك الفترة وذلك عند مستوى معنوية α.

وتقدر فترة الثقة لمعامل الانحدار β بالفترة التالية:

$$b - s_b t_{\upsilon,\alpha/2} \leq \beta \leq b + s_b t_{\upsilon,\alpha/2} \qquad (١٣,٢٦)$$

ثانيا: اختبار الفروض حول المعامل γ.

يصاغ الفرض الإحصائي كما يلي:

$$H_0 : \gamma_0 = 0$$

$$H_1 : \gamma \neq 0$$

ثم تحسب الإحصاءة t_0 حيث υ تمثل درجات الحرية وهي $\upsilon = n - 2$ والإحصاءة t_0 في هذه الحالة تحت صحة H_0 تعطى بالصيغة (١٣,٢٧) التالية:

$$t_0 = \frac{a - \gamma_0}{s_a} \qquad (١٣,٢٧)$$

حيث s_a تمثل الخطأ المعياري لمعامل a ويعطى بالعلاقة (١٣,٢٨) التالية:

$$s_a = s_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\Sigma x^2 - n\bar{x}^2}} \qquad (١٣,٢٨)$$

وتحدد منطقة قبول H_0 بالفترة $(-t_{\upsilon,\alpha/2}, t_{\upsilon,\alpha/2})$ حيث الاختبار البديل من طرفين ومنطقة رفض H_0 خارج تلك الفترة وذلك عند مستوى معنوية α.

وتقدر فترة الثقة للمعامل γ بالفترة (١٣,٢٩) التالية:

$$a - s_a t_{\upsilon,\alpha/2} \leq \gamma \leq a + s_a t_{\upsilon,\alpha/2} \qquad (١٣,٢٩)$$

ونوضح طريقة الحساب لما سبق من خلال المثال التالي:

مثال (١٣,١٣)

من مثال (١٣,٢) لدرجات الطلاب في مادة الرياضيات (y) والإحصاء (x) اختبر الفروض حول معامل الانحدار β والمعامل γ ثم أوجد فترة الثقة لكل من γ، β عند مستوى معنوية $\alpha = 0.05$.

الحل

سبق أن حسبنا الخطأ المعياري للتقدير في مثال (١٣,١١) السابق وكانت

قيمته هي $s_e = 1.55$.

كما سبق أن حسبنا في مثال (١٣,٢) القيم a ، b ، $\bar{x}$ و Σx^2 حيث كانت $\Sigma x^2 = 1398$ ، $\bar{x} = 12.75$ ، $b = 0.8744$ و $a = 0.977$ والكميات السابقة سوف نحتاجها لحساب الخطأين المعياريين s_b ، s_a المطلوبين لاتمام اختبار الفروض حول β و γ وتكوين فترة الثقة لهما عند مستوى معنوية $\alpha = 0.05$.

أولا: اختبار الفروض لمعامل الانحدار β ثم تقدير فترة الثقة له كما يلي:

ا) صياغة الفرض الإحصائي

$$H_0 : \beta_0 = 0$$

$$H_1 : \beta \neq 0$$

ب) اختيار الإحصاءة ثم حسابها

نختار الإحصاءة t_0 المعطاة بالعلاقة (١٣,٢٣) وهي:

$$t_0 = \frac{b - \beta_0}{s_b}$$

وتحسب s_b معطاة بالعلاقة (١٣,٢٥) كما يلي:

$$s_b = \frac{s_e}{\sqrt{\Sigma x^2 - n\bar{x}^2}} = \frac{1.55}{\sqrt{1398 - 8(12.75)^2}} = \frac{1.55}{9.874} = 0.157$$

وحيث $\upsilon = 8 - 2 = 6$ فإن t_0 تحسب من العلاقة السابقة كما يلي:

$$t_0 = \frac{b - \beta_0}{s_b} = \frac{0.8744 - 0}{0.157} = 5.569$$

ج) تحديد منطقة قبول ورفض H_0 عند مستوى معنوية $\alpha = 0.05$

حيث تكون منطقة قبول H_0 هي $(-t_{6,0.025} , t_{6,0.025})$ باستخدام جدول t رقم (٢) آخر الكتاب تصبح فترة قبـول H_0 هي $(-2.45, 2.45)$ ومنطقة رفض H_0 خارج هذه الفترة.

د) القرار الإحصائي

حيث $t_0=5.569$ المحسوبة واقعة خارج فترة قبـــول H_0 وهي $(-2.45, 2.45)$ فإننا نرفـض H_0 أي أن معامل الانحـدار β معنوي ويختلف عن الصفر عـنـد مستوى معنوية $\alpha = 0.05$ وتقدر فترة الثقة لمعامل الانحدار β باستخدام فترة الثقة المعطاة بالصيغة (١٣,٢٦) السابقة كما يلي:

$$b - s_b \ t_{6,\,0.025} \ \leq \beta \leq b + s_b \ t_{6,\,0.025}$$

$$0.8744 - 0.157(2.45) \leq \beta \leq 0.8744 + 0.157(2.45)$$

$$0.49 \leq \beta \leq 1.26$$

أي أن تقدير الحد الأعلى لمعامل الانحـدار β لفترة الثقة هـو 1.26 وأن تقدير الحـد الأدنى لمعامل الانحدار β لفترة الثقة هو 0.49 وذلك بدرجة ثقة قدرها 95٪.

ثانيا: اختبار الفروض حول المعامل γ ثم تقدير فترة الثقة له كما يلي:

ا) صياغة الفرض الإحصائي

$$H_0 : \gamma_0 = 0$$

$$H_1 : \gamma \neq 0$$

ب) اختيار الإحصاءة ثم حسابها

نختار الإحصاءة t_0 المعطاة بالعلاقة (١٣,٢٦) وهي:

$$t_0 = \frac{a - \gamma_0}{s_b} = \frac{a - 0}{s_b}$$

وتحسب s_a المعطاة بالعلاقة (١٣,٢٨) السابقة كما يلي:

$$s_a = s_e \sqrt{\frac{1}{n} + \frac{\overline{x}^2}{\Sigma x^2 - n\,\overline{x}^2}} = 1.55 \sqrt{\frac{1}{8} + \frac{(12.75)^2}{1398\text{-}8(12.75)^2}} = 2.0751$$

وحيث $\upsilon = 8 - 2 = 6$ فإن الإحصاءة t_0 تحسب من العلاقة السابقة كما يلي:

$$t_0 = \frac{a - \gamma_0}{s_a} = \frac{0.98 - 0}{2.0751} = 0.4723$$

ج) تحديد منطقة قبول ورفض H_0 عند مستوى معنوية $\alpha = 0.05$

حيث تكون منطقة قبول H_0 هي$(-t_{6,0.025}\ ,\ t_{6,0.025})$ ومن جدول t رقم (٢) آخر الكتاب تصبح هذه الفترة هـي $(-2.45, 2.45)$. ومنطقة رفض H_0 خارج هذه الفترة.

د) القرار الإحصائي

حيث $t_0 = 0.4723$ المحسوبة واقعة داخل فترة قبـول H_0 وهي $(-2.45, 2.45)$ فإننا لا نستطيع رفض H_0 أي أن المعامل γ يساوي صفرا أي أن خط الانحدار يمر بنقطة الأصل.

ولتقدير فترة الثقة للمعـامـل γ باستخدام فترة الثقة المعطاة بالصيغـة (١٣,٢٩) السابقة يتم كما يلي.

$$a - s_a t_{6,\ 0.025} \le \gamma \le a + s_a t_{6,\ 0.025}$$

$$0.98 - 2.07519\ (2.45) \le \gamma \le 0.98 + 2.0751(2.45)$$

$$-4.1 \le \gamma \le 6.1$$

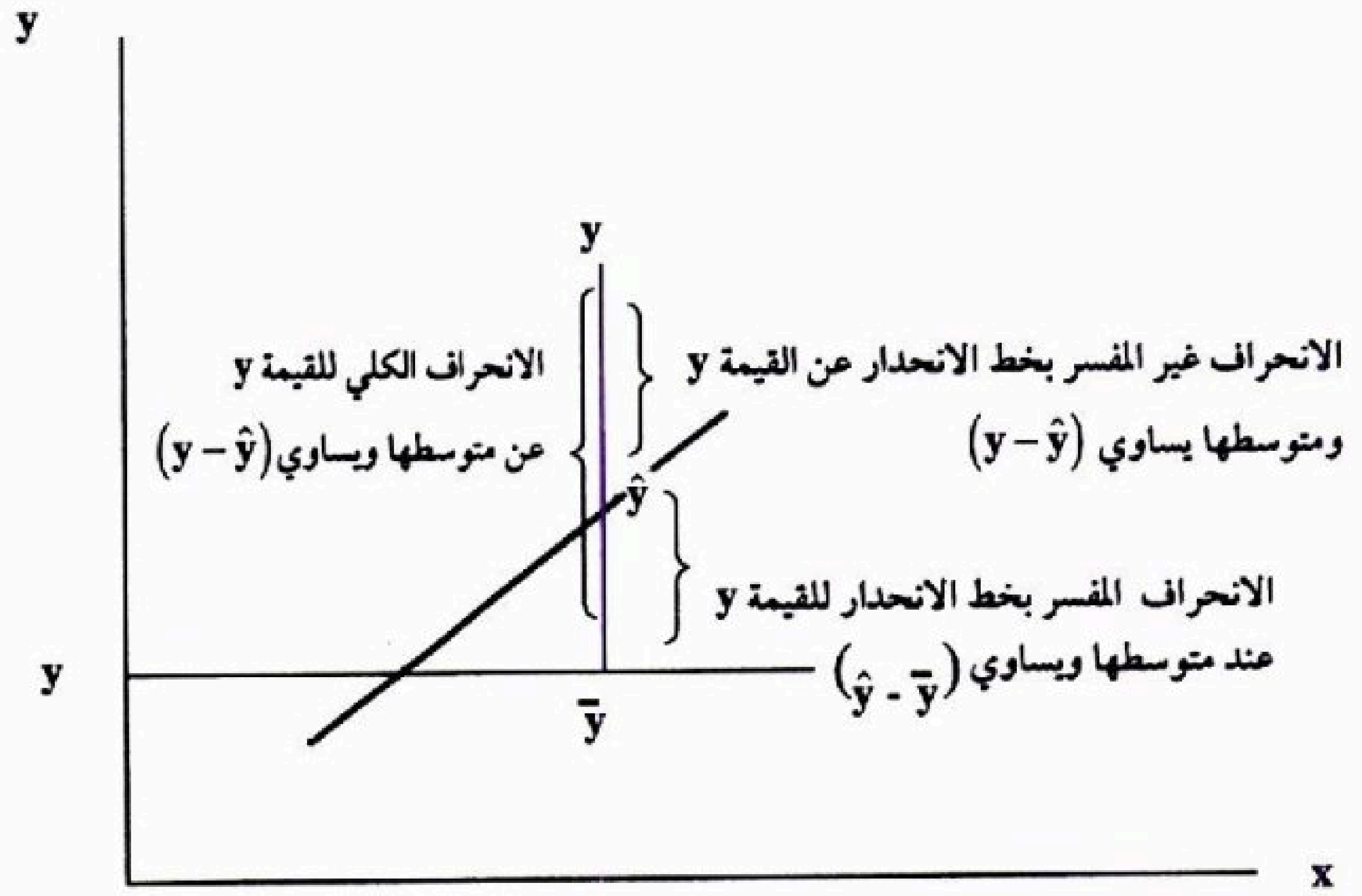
أي أن تقدير الحد الأعلى لفترة الثقة للمعامـل $\gamma = 6.1$ وتقدير الحد الأدنى لفترة الثقة للمعامل $\gamma = -4.1$.

ونلاحظ أن هذه الفترة تحوي صفرا أي أنه احتمال 95٪ قد تكون γ قريبة جدا من الصفر.

(١٠ , ١٣) معامل التحديد

Coefficent of Determination

يعرف معامل التحديد بمقدار التغير في y الذي تفسره معادلة خط الانحدار (٢١ , ١٣) السابقة، ونوضح ذلك بالشكل (١٣ , ١٣) كما يلي:



شكل (١٣ , ١٣)

حيث نعتبر قيمة واحدة من المشاهدات y كما هي موضحة في شكل (١٣ , ١٣). الآن، إذا استخدمنا المشاهدات $\bar{y}$ للتقدير فإن الانحراف الكلي لهذه القيمة y عن متوسطها $\bar{y}$ هو $(y - \bar{y})$. الآن لو استخدمنا خط الانحدار لتقدير y فإننا سنحصل على تقدير أفضل وفي هذه الحالة فإن خط الانحدار سوف يفسر فقط $(\hat{y} - \bar{y})$ قِبَلْ الانحراف الكلي $(y - \bar{y})$ ويبقى الجزء $(y - \hat{y})$ من الانحراف الكلي غير مفسر. الآن لنعتبر مجموعة جميع المشاهدات في العينة وليس مشاهدة واحدة فقط، التغير الكلي والذي هو مجموع مربعات الانحرافات لهذه المشاهدات عن متوسطها

هو $\Sigma(y - \overline{y})^2$، والجزء المفسر من هذا التغير الكلي أو مجموع مربعات الانحرافات المفسرة لهذه المشاهدات عن متوسطها هو $\Sigma(\hat{y} - \overline{y})^2$، والجزء غير المفسر من هذا التغير الكلي أو مجموع مربعات الانحرافات غير المفسر لهذه المشاهدات من خط الانحدار هو $\Sigma(y - \hat{y})^2$، وبالتالي فإن مجموع مربعات الانحرافات عن المتوسط = مجموع مربعات الانحرافات المفسرة عن المتوسط + مجموع مربعات الانحرافات غير المفسر عن المتوسط ويكتب الرموز بالصيغة (١٣,٣٠) التالية:

$$\Sigma(y - \overline{y})^2 = \Sigma(\hat{y} - \overline{y})^2 + \Sigma(y - \hat{y})^2 \qquad (١٣,٣٠)$$

وبقسمة العلاقة (١٣,٣٠) على المقدار $\Sigma(y - \overline{y})^2$ نحصل على:

$$1 = \frac{\Sigma(\hat{y} - \overline{y})^2}{\Sigma(y - \overline{y})^2} + \frac{\Sigma(y - \hat{y})^2}{\Sigma(y - \overline{y})^2} \qquad (١٣,٣١)$$

والكمية $\frac{\Sigma(\hat{y} - \overline{y})^2}{\Sigma(y - \overline{y})^2}$ المعطاة بالحد الأول في الطرف الأيمن للعلاقة (١٣,٣١) هي عبارة عن الجزء المفسر بمعادلة خط الانحدار للتغير الكلي ويسمى بمعامل التحديد ويرمز له بالرمز r^2 وبالتعويض عن الكمية $\frac{\Sigma(\hat{y} - \overline{y})^2}{\Sigma(y - \overline{y})^2}$ بالمقدار r^2 في العلاقة (١٣,٣١) السابقة نحصل على:

$$1 = r^2 + \frac{\Sigma(y - \hat{y})^2}{\Sigma(y - \overline{y})^2}$$

أي أن

$$r^2 = 1 - \frac{\Sigma(y - \hat{y})^2}{\Sigma(y - \bar{y})^2} \qquad (١٣,٣٢)$$

والقيمة r^2 المعطاة بالصيغة (١٣,٣٢) تقيس مدى مقدرة المتغير x في تفسير المتغير y.
ويمكن تبسيط العلاقة (١٣,٣٢) أكثر لايجاد قيمة r^2، وذلك باستخدام العلاقة (١٣,٢٠) السابقة للبسط والقيمة $SS(y) = \Sigma y^2 - n\bar{y}^2$ للمقام فتصبح العلاقة (١٣,٣٢) كما يلي:

$$r^2 = 1 - \frac{(\Sigma y^2 - a\Sigma y - b\Sigma xy)}{\Sigma y^2 - n\,\bar{y}^2}$$

أي أن

$$r^2 = \frac{a\Sigma y + b\Sigma xy - n\bar{y}}{\Sigma y^2 - n\bar{y}^2} = \frac{a\Sigma y + b\Sigma xy - n\bar{y}}{SS(y)}$$

أي أن

$$r^2 = \frac{a\Sigma y + b\Sigma xy - n\bar{y}^2}{\Sigma y^2 - n\bar{y}^2} \qquad (١٣,٣٣)$$

ملاحظة

العلاقة (١٣,٣٣) أسهل بكثير في حساب r^2 من العلاقة (١٣,٣٢) السابقة.

مثال (١٣,١٤)

أوجد معامل التحديد r^2 لانحدار درجات الطلاب في الرياضيات (y) والإحصاء (x) في مثال (١٣,٢) السابق.

الحل

من جدول (١٣,٢) نحصل على المجاميع التالية:

$$\Sigma y = 97,\ \Sigma xy = 1322,\ \Sigma y^2 = 1265,\ \overline{y} = 12.125$$

من مثال (٩ ,١٣) حيث وجدنا معادلة خط الانحدار فحصلنا على $a = 0.977, b = 0.8744$ وبالتعويض في الصيغة (٣٣ ,١٣) نحصل على معامل التحديد r^2 كما يلي:

$$r^2 = \frac{a\Sigma y + b\Sigma xy - n\overline{y}}{\Sigma y^2 - n\overline{y}^2}$$

$$= \frac{(0.977)(97) + (0.8744)(1322) - 8(12.125)^2}{1265 - 8(12.125)^2}$$

$$= \frac{74.6008}{88.872} = 0.8393 = 0.84$$

أي أن 84 ٪ من التغير في y أمكن تفسيره بواسطة x.

ملاحظة

مقدار معامل الارتباط الخطي لبرسون $|r|$ هو الجذر التربيعي لمعامل التحديد r^2 (يعطي مقدار الارتباط ولكن لا يعطي اتجاهه) أي أن

$$|r| = \sqrt{r^2}$$

ومن مثال (١٤ ,١٣) السابق فإن:

$$|r| = \sqrt{0.3893} = 0.916133$$

ويمكن معرفة اتجاه r أو إشارته بمعرفة إشارة معامل الانحدار b أي أن:

$$r = \text{sign}(b)\sqrt{r^2}$$

حيث sign(b) هي إشارة b فمثلا $\text{sign}(4) = +1,\ \text{sign}(-4) = -1$.

(١١ , ١٣) معادلات انحدار غير خطية ويمكن تحويلها الى معادلات خطية
Non Linear Regressions Can be Transform to Linear Regressions

المعادلة

$$y = a + b x$$

تسمى معادلة انحدار خطي وذلك لأن المعامــلات a, b تظهر خطيا في هذه المعادلة أي بمعنى أنه إذا اشتققنا الطرف الأيمن لهذه المعادلة جزئيا بالنسبة للمعاملات a,b فإن المعادلات الناتجة لا تعتمد على قيم a,b أي أن:

$$\frac{\partial}{\partial a}(a + b x) = 1$$

$$\frac{\partial}{\partial b}(a + b x) = x$$

وهما لا يعتمدان على a أو b بعكس المعادلة التالية

$$y = a e^{bx} \qquad (١٣,٣٤)$$

فإن اشتقاق الطرف الأيمن للمعادلة (١٣,٣٤) بالنسبة للمعــامــلات b, a فنحصــل على:

$$\frac{\partial}{\partial a}(a e^{bx}) = e^{bx}$$

$$\frac{\partial}{\partial b}(a e^{bx}) = a x e^{bx}$$

وكلاهما يعتمد على a و b.

المعادلة (١٣,٣٤) تسمى معادلة انحدار خطي وكذلك المعادلات التالية:

$$y = a\,x^{b} \tag{١٣,٣٥}$$

$$y = a + e^{b}\,x \tag{١٣,٣٦}$$

$$y = \frac{a}{1 + b\,e^{cx}} \tag{١٣,٣٧}$$

وهذه فقط أمثلة من كثير. والمعادلات (١٣,٣٤)، (١٣,٣٥) و (١٣,٣٦) يمكن تحويلها إلى معادلات خطية ومن ثم دراستها بالطرق السابقة في هذا الفصل. أمـا المعادلة (١٣,٣٧) لا يمكن تحويلها إلى معادلة خطية ولهذا فإن دراستها خارج نطاق هذا الكتاب.

ولتحويل معادلات الانحدار غير الخطي و التي يمكن تحويلها إلى معـادلات خطية:

$$y' = \ln y \qquad a' = \ln a$$

وبأخذ اللوغاريتم الطبيعي لكلا طرفي (١٣,٣٤) نجد أن

$$\ln y = \ln a + b\,x \tag{١٣,٣٨}$$

ويمكن وضع الصيغة (١٣,٣٨) كما يلي:

$$y' = a' + b\,x \tag{١٣,٣٩}$$

الآن المعادلة (١٣,٣٩) على الشكل الخطي السابق.

وبنفس الطريقة المعادلة (١٣,٣٥) ويأخذ لوغاريتم الطرفين نجد أن:

$$\ln y = \ln a + b \ln x \tag{١٣,٤٠}$$

وبوضع القيم في العلاقة (١٣,٤٠) بالرموز التالية: $y' = \ln y\ ,\ a' = \ln a\ ,\ x' = \ln x$

نجد أن الصيغة (١٣,٤٠) اصبحت بالصيغة التالية:

$$y' = a' + b\,x' \qquad (١٣,٤١)$$

ونجد الصيغة (١٣,٤١) على الشكل الخطي السابق. كذلك بالنسبة للصيغة (١٣,٣٦) بالتعويض عن $b' = e^{b}$ نجد أن:

$$y = a + b'\,x \qquad (١٣,٤٢)$$

ونجد أن الصيغة (١٣,٤٢) على الشكل الخطي السابق. طبعا بعد إجراء التحليل المناسب بالطرق التي تم دراستها في هذا الفصل للانحدار الخطي يعاد تحويل المتغيرات والعوامل بأخذ التحويل العكسي للوغاريتم حتى نعود إلى طبيعة الشكل الأساسي للعلاقة محل الدراسة.

مثال (١٣,١٥)

البيانات التي بالجدول التالي فيه x تمثل فترة باليوم لتعاطي مريض لمضاد حيوي (لمرض التهاب بكتيري حاد) و y تمثل النسبة المئوية لنكسة تصيب المريض.

x	2	5	7	10	14	19	26	31	34	38
y	54	50	45	37	35	25	20	16	18	13
ln y	3.989	3.912	3.807	3.611	3.555	3.219	2.996	2.773	2.890	2.565

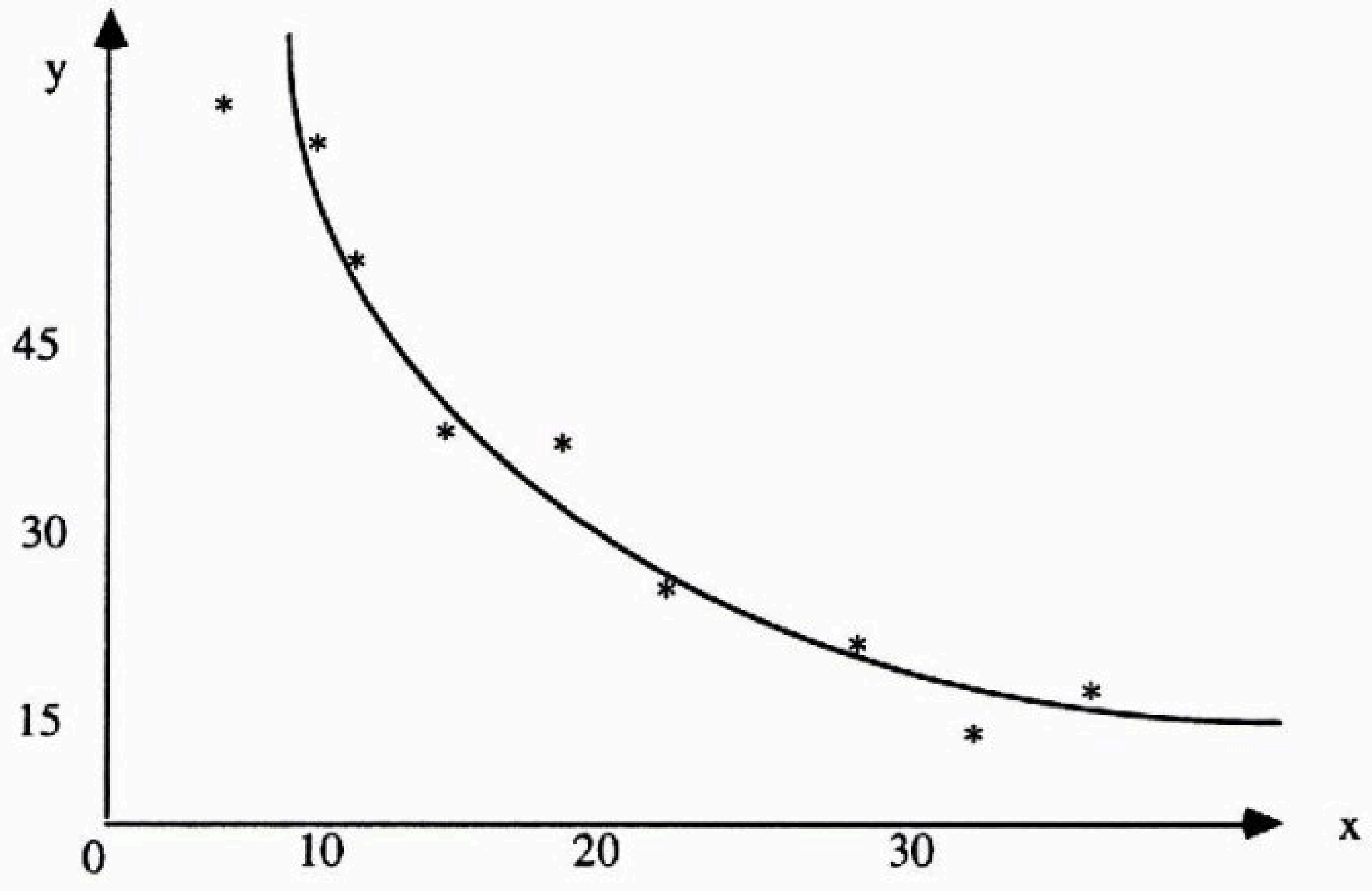
المطلوب:

(i) ارسم مخطط انتشار كل من (x,y) و (x, ln y) ثم استنتج أيهما يكون مناسبا لتقدير معادلة انحدار y على x.

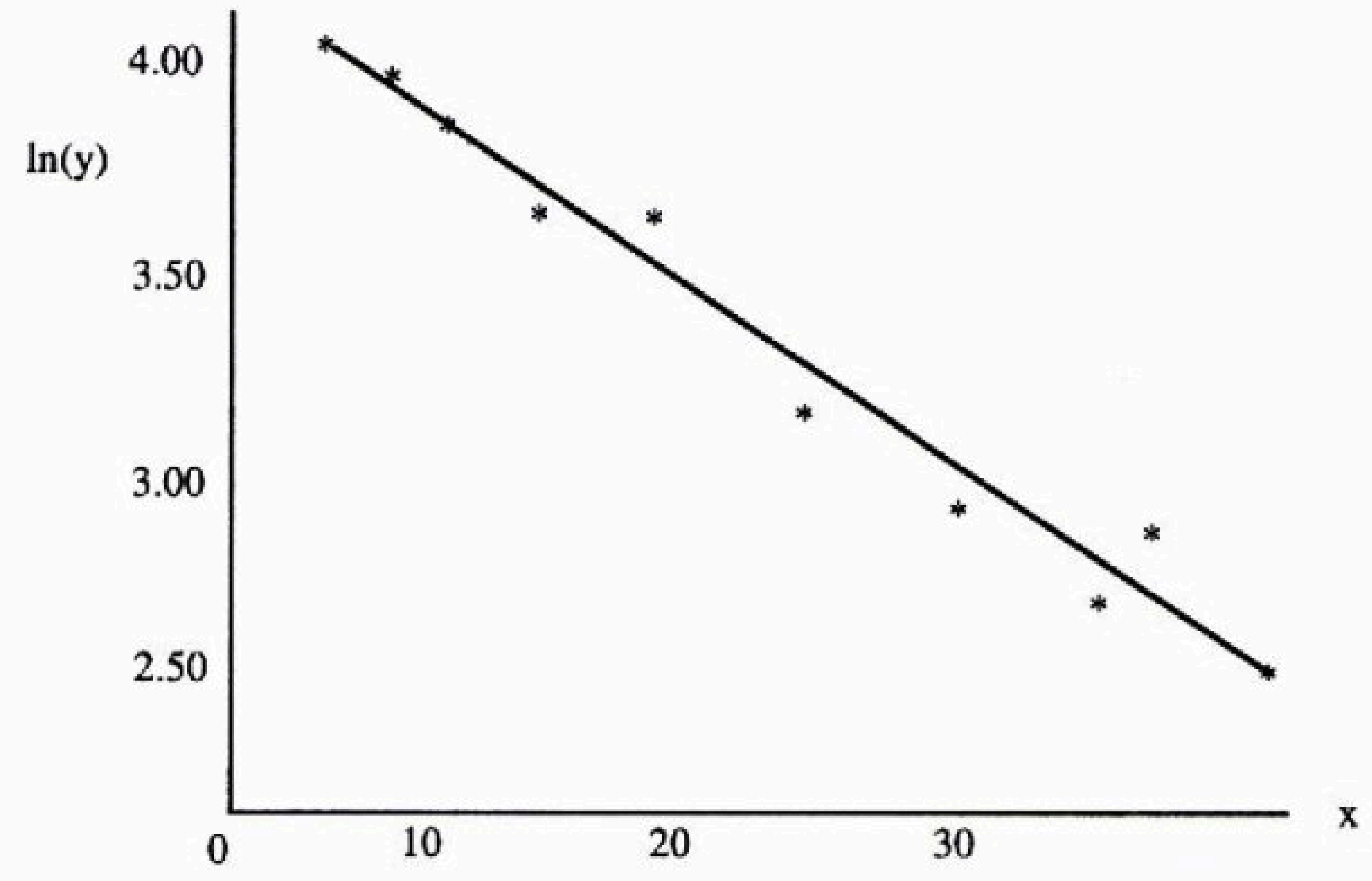
(ii) قدّر معادلة انحدار y على x ثم احسب معامل التحديد r^2 للنموذج المقترح.

الحل

نرسم مخططي الانتشار، شكل (١٣,١٤) يمثل (x,y)، وشكل (١٣,١٥) يمثل (x, ln y) من البيانات التي بالمثال (١٣,١٥) كما يلي:



شكل (١٤, ١٣)



شكل (١٥, ١٣)

من مخططي الانتشار شكلي (١٣,١٤) و (١٣,١٥)، نجد أن الشكـل (١٣,١٥) يكون مناسبا أكثر لتقدير معادلة خط انحـدار x, ln y والمعطاة بالصيغـة (١٣,٣٩) السابقة وهي:

$$y' = a' + b\,x$$

حيث $y' = \ln y \quad , \quad a' = \ln a$

ولتسهيل الحل نكون جدول (١٣,١٠) كما يلي:

جدول (١٣,١٠)

Y´ = ln y	X	XY´	X ²	Y´ ²
3.989	2	7.978	4	15.912
3.912	5	19.56	25	15.304
3.807	7	26.649	49	14.493
3.611	10	36.11	100	13.039
3.555	14	49.77	196	12.638
3.912	19	61.161	361	10.362
2.996	26	77.896	676	8.976
2.773	31	85.963	961	7.690
2.890	34	98.26	1156	8.360
2.563	38	97.394	1444	6.570
33.315	186	463.347	4972	113.344

من الجدول (١٣,١٠) نحسب a´ و b كما يلي:

$$b = \frac{n\Sigma xy' - \Sigma x\, \Sigma y'}{n\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{10(463.347) - 186(33.315)}{10(4972) - (186)^2} = -0.0389$$

$$a' = \overline{Y}' - b\overline{X}$$

$$= 3.3315 - (-0.0380)(18.6) = 4.0558$$

وبذلك تصبح معادلة انحدار $\ln y$ على x هي

$$\ln y = 4.0558 - 0.0389\, x$$

ولحساب معامل التحديد r^2 من الصيغة (١٣,٣٣) وهي

$$r^2 = \frac{a'\Sigma y' + b\Sigma xy' - n\overline{y'}}{\Sigma y'^2 - n\overline{y'}^2}$$

$$= \frac{(4.0558)(33.315) + (-0.0389)(463.347) - 10(3.332)}{113.344 - 10(3.332)^2} = 0.98$$

أي أن معامل التحديد يفسر 98% من التغير في y بواسطة x. وباستخدام التحويل العكسي للوغاريتم نحصل على نموذج انحدار y على x كما يلي

$$y = 57.73e^{-0.039x}$$

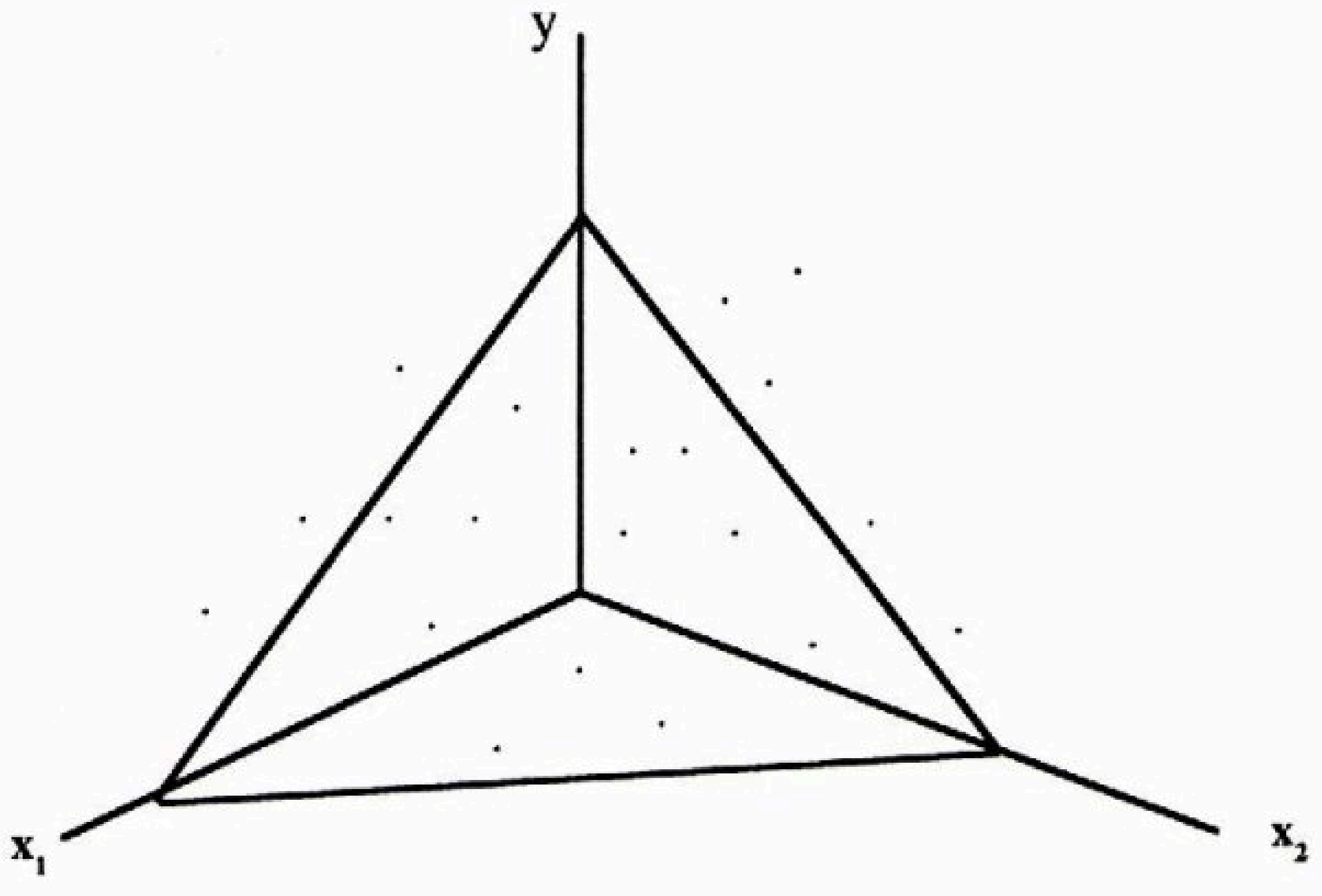
ملاحظة

يمكن اختبار الفروض للعوامل γ, β وتقدير فترة الثقة لهما عند مستوى معنوية α مثل ما تم بالنسبة لنموذج خط الانحدار في بند (١٣,٨) السابق حيث يتم حساب الخطأ المعياري للتقدير s_e بالمقدار $s_e = \sqrt{\frac{\Sigma(y_i - \hat{y})^2}{n-2}}$.

(١٣,١٢) الانحدار الخطي المتعدد

Multiple Linear Regression

في أغلب مسائل التنبؤ العملية نجد أن عملية التنبؤ تعتمد على أكثر من متغير. مثال ذلك: إذا أرادت إحدى الكليات تطوير الصيغة المستخدمة في عملية القبول معتمدة على التنبؤ بمدى النجاح الأكاديمي للطالب الجديد المقبول، وترغب عند تطوير هذه الصيغة الأخذ في الاعتبار بعض المتغيرات مثل متوسط درجات الطالب في الثانوية العامة، ومتوسط درجات اختبار القدرات، ومتوسط درجات الاختبار الشخصي، وإدخال هذه المتغيرات الثلاث في دالة الانحدار. كذلك عند التنبؤ باستهلاك الكهرباء لعشرة أعوام قادمة في مدينة كبيرة، فإن مخططي المدينة يأخذوا في الاعتبار متغيرات كثيرة لها صلة باستهلاك الكهرباء منها النمو الصناعي، والاستهلاك المتزايد داخل المنازل وغيرها. والطرق المستخدمة للتنبؤ بقيم متغير بدلالة متغيرات أخرى مشابهة للطرق المستخدمة في التنبؤ بقيم متغير بدلالة متغير واحد (الانحدار الخطي البسيط). على سبيل المثال: إذا أردنا تحديد المتغير y بدلالة دالة خطية في متغيرين x_1, x_2 فإن المسألة تؤول إلى حساب أفضل سطح يمر بنقط الشكل الانتشاري في الأبعاد الثلاثة كما هو موضح في شكل (١٣,١٦) كما يلي: وذلك باستخدام طريقة المربعات الصغرى.



شكل (١٦ , ١٣)

وحيث إن معادلة سطح في ثلاثة أبعاد باستخدام المتغيرات y, x_1, x_2 يمكن كتابتها بالصيغة (٤٣ , ١٣) كما يلي:

$$y = a + b_1 x_1 + b_2 x_2 \qquad (١٣ , ٤٣)$$

فإن المشكلة في الصيغة (٤٣ , ١٣) تؤول إلى حساب المعالم الـثـلاث a, b_1, b_2 بطريقة المربعات الصغرى، ذلك بحل مجموعة من ثلاث معادلات وقد فرضنا أن المجموع قد تم على قيمة n من المتغيرات وتصبح المعادلات كما يلي:

$$\begin{aligned} an + b_1 \Sigma x_1 + b_2 \Sigma x_2 &= \Sigma y \\ a \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 &= \Sigma x_1 y \\ a \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2 &= \Sigma x_2 y \end{aligned} \qquad (١٣ , ٤٤)$$

وهذا الأسلوب لطريقة المربعات الصغرى يمكن تعميمه لأي متغيرات إضافية.
ونوضح طريقة الحساب من خلال المثال التالي:

مثال (١٣,١٦)

في عينة مكونة من 8 منازل خاصة بالأسر كان سعر بيعها (y) معتمدا على عدد غرف النوم وعدد الحمامات بالمنزل وكانت البيانات معطاة بالألف ريال كما يلي:

y ثمن المنزل بالريال	338	293	388	292	347	299	434	379
x_1 عدد غرف النوم بالمنزل	3	2	4	2	3	2	5	4
x_2 عدد الحمامات للمنزل	2	1	3	1	2	2	3	2

والمطلوب ايجاد معادلة انحدار ثمن المنزل y على عدد غرف النوم x_1 وعدد الحمامات x_2، ثم تقدير ثمن المنزل y عندما يكون المنزل يشتمل على ثلاثة غرف نوم وعدد 2 حمام أي أن($x_1 = 3 \ , \ x_2 = 2$).

الحل

بناء على البيانات الواردة بهذا المثال فإننا نستخدم النموذج المعطى بالصيغة (١٣,٤٣) وهي $y = a + b_1 x_1 + b_2 x_2$. ولتسهيل الحل نكون جدول (١٣,١١) كما يلي:

بالتعويض بقيم المجاميع من الجدول (١٣,١١) في المعادلات رقم (١٣,٤٤) حيث

$$\Sigma x_1 y = 9061 \ , \ \Sigma y = 2770 \ , \ \Sigma x_2 = 16 \ , \ \Sigma x_1 = 25 \ , \ n = 8$$

$$\Sigma x_2^2 = 36 \ , \ \Sigma x_1 x_2 = 55 \ , \ \Sigma x_1^2 = 87 \ , \ \Sigma x_2 y = 5777$$

جدول (١١ ,١٣)

	x_1	x_2	y	x_1y	x_2y	x_1^2	x_1x_2	x_2^2
	3	2	338	1014	766	9	6	4
	2	1	293	586	293	4	2	1
	4	3	388	1552	1164	16	12	9
	2	1	292	584	292	4	2	1
	2	2	3.7	1041	964	9	6	4
	2	3	299	598	598	4	4	4
	5	3	434	2170	1302	25	15	9
	4	2	379	1516	758	16	8	4
المجموع	25	16	2770	9061	5777	87	55	36

نحصل على المعادلات التالية:

$$8\,a + 25b_1 + 16b_2 = 2770$$

$$25\,a + 87b_1 + 55b_2 = 9011$$

$$16\,a + 55b_1 + 36b_2 = 5777$$

وبحل المعادلات السابقة نحصل على قيم العوامل a, b_1, b_2 حيث $a = 201.97$, $b_1 = 41.49$, $b_2 = 73.10$ وتصبح معادلة خط الانحدار المتعدد بالتعويض بقيم a, b_1, b_2 السابقة في الصيغة (١٣ ,٤٣) كما يلي:

$$y = 201.97 + 41.49\,x_1 + 73.10\,x_2$$

ولتقدير ثمن المنزل الذي يحتوي على ثلاث غرف نوم$(x_1 = 3)$ وعلى عدد اثنين

من الحمامات($x_2 = 2$) هو بالتعريض بقيم x_2 , x_1 في الصيغة السابقة نحصل على y ثمن المنزل كما يلي:

$y = 201.97 + 41.49(4) + 7310(2) = 341.06$ الف ريال

ملاحظة

يمكن حساب الخطأ المعياري s_e بالعلاقة $s_e = \sqrt{\dfrac{\Sigma(y_i - \hat{y}_i)^2}{n-3}}$.

(١٣,١٣) تمارين

١ - ما هي الصور المختلفة لمعاملات الارتباط. اشرح كيف يمكن تفسير اتجاه وقوة معامل الارتباط.

٢ - سحب أحد الباحثين عينة من عشرة أزواج من ملاحظاته فيما يختص بطـول الشخص (y) وعمره (x) فكانت البيانات كما يلي:

الطول (y)	170	171	167	160	164	180	182	168	170	162
العمـر (x)	20	22	24	20	25	30	32	22	22	25

المطلوب:

(i) ارسم مخطط انتشار بين x , y.

(ii) احسب معامل الارتباط لبرسون بين الطول والعمر.

(iii) احسب معامل الارتباط للرتب بين الطول والعمر.

(iv) اختبر معنوية الارتباط لكل من معاملي الارتباط لبيرسون وللرتب عند مستوى معنوية $\alpha = 0.01$.

٣ - الجدول التالي يمثل تقديرات ثمانية طلاب في مادتي الإحصاء والفيزياء

تقديرات الإحصاء	E	B	B	C	D	D	E	A
تقديرات الفيزياء	D	A	B	E	D	C	E	B

(i) أوجد معامل الارتباط بين تقديرات الإحصاء والفيزياء.

(ii) اختبر معنوية هذا المعامل عندما $\alpha = 0.05$.

٤ - الجدول التالي يمثل تكاليف الدعاية بمئات الريالات x والمبيعات y بمئات الريالات أيضا:

تكاليف الدعاية x	7	5	4	10	9	10	9
المبيعات y	140	120	120	190	150	180	160

(i) ارسم مخطط انتشار للمتغيرين x , y.

(ii) ارسم معامل الارتباط بين تكاليف الدعاية والمبيعات.

(iii) اختبر معنوية معامل الارتباط عند مستوى معنوية $\alpha = 0.05$.

٥ - الجدول التالي يمثل عمر الزوج x وعمر الزوجة y لعينة مكونة من عشرة أسر:

عمر الزوج x	70	55	52	51	28	38	40	29	25	60
عمر الزوجة y	65	55	50	38	20	30	40	21	17	60

(i) ارسم مخطط انتشار المتغيرين x, y.

(ii) احسب معامل الارتباط بطريقتين مختلفتين.

(iii) اختبر معنوية كل من معاملي الارتباط عند مستوى معنوية $\alpha = 0.01$.

٦ - في تجربة لدراسة تأثير تطعيم مجموعة من الحيوانات ضد مرض معين كانت

النتائج كما يلي:

التطعيم \ الإصابة	لم يصب بالمرض	أصيب بالمرض
طعم	12	5
لم يطعم	4	9

(i) احسب معامل الاقتران بين التطعيم والإصابة بالمرض.

(ii) اوضح مدى تأثير التطعيم في الوقاية من هذا المرض.

٧ - كانت نتيجة دراسة ألوان البشرة لمجموعة من الأمهات مع أول مولود لهن كما يلي:

أول مولود \ الأمهات	أبيض	قمحي	أسمر
أبيض	27	6	7
قمحي	8	17	5
أسمر	5	7	18

بيّن فيما إذا كان هناك توافق بين لون بشرة الطفل الأول والأم.

٨ - أجري بحث في إحدى عيادات العلاج النفسي عن مدى ارتباط الوضع الاجتماعي ونوعية المرض فكانت النتائج كما يلي:

نوع المرض / الوضع الاجتماعي	فصام شخصية	اضطرابات شخصية	كآبة	أعصاب
عـــال	8	12	5	25
متوسط	12	15	20	8
منخفض	9	8	11	7

ادرس التوافق بين الوضع الاجتماعي ونوع المرض.

٩ - (i) ما هو المقصود بانحدار y على x.

(ii) كيف يستطيع الباحث تحديد درجة معادلة الانحدار.

(iii) استعرض باختصار الصورالمختلفة للمعادلات الانحدارية.

(iv) ما هو الفرق بين الانحدار البسيط والانحدار المتعدد.

(v) كيف يمكن قياس الخطأ الناتج من وجود انحرافات بين القيم الأصليـة والقيم المقدرة باستخدام معادلة انحدار y على x.

١٠ - لنفرض أن x ترمز إلى درجات الحرارة المئوية بين الساعة 3.00 والساعة 4.00 بعد الظهر خلال فصل الصيف، y ترمز إلى كمية الطاقة الكهربائية المستهلكة ممثلة بمستويات تدرج من 1 إلى 10 حيث تمثل 10 أعلى استهلاك. وتم تسجيل البيانات الآتية خلال فترة طولها 10 أيام

درجة الحرارة X	21	31	30	25	34	30	32	32	38	38
كمية الطاقة المستهلكة Y	4	7	6	5	8	7	4.5	6	9	9.5

(i) ارسم شكل انتشار لهذه البيانات لتتأكد من إمكانية تمثيلها بخط مستقيم.

(ii) قدّر معادلة خط انحـدار y على x واستخدمها في تقدير الاستهلاك مـن

الطاقة الكهربائية عندما تكون درجة الحرارة °35.

١١ - لوحظت مجموعة من نباتات القمح الصغيرة لفترة طولها سبعة أيام عند مستويات محددة ومختلفة من الرطوبة x فكانت نتائج نموها y بالبوصة كما يلي:

مستويات الرطوبة x	75	70	65	60	55	55	50	50	45	40
نمو النبات y	4.0	3.9	3.7	4.2	3.0	2.8	2.2	1.8	1.2	1.0

(i) أوجد معادلة خط انحدار y على x.

(ii) احسب الخطأ المعياري للتقدير s_e.

(iii) أوجد تقدير فترة الثقة لمعامل الانحدار β عند مستوى معنوية $\alpha = 0.05$.

(iv) اختبر معنوية معامل الانحدار β عند مستوى معنوية $\alpha = 0.05$.

(v) اختبر معنوية المعامل γ وتقدير فترة الثقة له عند مستوى معنوية $\alpha = 0.05$.

١٢ - فيما يلي البيانات الخاصة بأعمار x أحد الأنواع من الأشجار بالسنوات وأطوال y هذه الأشجار بالقدم.

العمر بالسنوات x	3	1	2	5	4
الطول بالقدم y	9	5	7	14	10

(i) أوجد قيمة الاختلاف الكلي.

(ii) أوجد الاختلاف المفسر وغير المفسر وتحقق من أن الاختلاف الكلي = الاختلاف المفسر + الاختلاف غير المفسر.

(iii) استنتج معامل التحديد من العلاقة السابقة.

(iv) استنتج معامل الارتباط من معامل التحديد ومقارنته بمعامل الارتباط الخطي لبيرسون بعد حسابه.

١٣ - الجدول التالي يبين عدد البكتيريا y الموجودة في وحدة الحجم في مزرعة للبكتيريا بعد x ساعة:

عدد الساعات x	0	1	2	3	4	5	6
عدد البكتيريا في وحدة الحجم y	32	47	65	92	132	190	275

(i) ارسم شكل الانتشار.

(ii) إذا كانت العلاقة بين المتغيرين (x, y) من النوع $y = ab^x$. قدّر قيم الثوابت a, b واستخدم المعادلة في تقدير عدد البكتيريا في وحدة الحجم بعد سبع ساعات ثم أوجد الخطأ في تقدير عدد البكتيريا.

١٤ - إذا كان من المعروف أن هناك علاقة بين فترة الإصابة بمرض معين وعدد البكتيريا في العضو المصاب ودرجة حرارة المريض عند دخوله المستشفى. تم اختيار 10 مصابين بهذا المرض، وتم عد البكتيريا لديهم ودرجة حرارتهم وقت دخول المستشفى، وسجلت أطوال فترات اصابتهم بالمرض، فحصلنا على النتائج التالية:

عدد البكتيريا بالآلاف x_1	8	7	4	6	9	7	8	3	8	7
درجة الحرارة المئوية x_2	37.6	39.2	38.5	37.4	38.1	39.1	39.0	37.8	38.2	39.1
فترة الإصابة باليوم y	11	10	11	8	12	11	13	7	10	11

(i) أوجد معادلة انحدار فترة الإصابة بالمرض على عدد البكتيريا ودرجة الحرارة.

(ii) أوجد الخطأ المعياري للتقدير s_e .

١٥ - الجدول التالي يمثل الدخل x والإنفاق y لمجموعة من الأسر بمئات الريالات

x الدخل	56	66	42	44	38	27	39	40
y الإنفـــــاق	31	38	27	22	19	25	20	28

(i) أوجد معادلة خط انحدار y على x .

(ii) أوجد معامل ارتباط بيرسون وسبيرمان للدخل والإنفاق

(iii) أوجد قيمة الإنفاق عندما يكون الدخل 6000 ريال .

(iv) أوجد تقدير فترة الثقة لمعامل الانحدار β عند مستوى معنوية $\alpha = 0.05$.

(v) اختبر معنوية معامل الانحدار β عنـد مستـوى معنويـة $\alpha = 0.05$.

١٦ - الجدول التالي يوضح السن من x وضغط الدم y لثمان من الإناث:

x السن	42	36	63	55	42	60	49	68
y ضغط الدم	125	118	140	150	140	155	145	152

(i) أوجد معامل ارتباط بين y و x

(ii) أوجد معادلة خط انحدار y على x .

(iii) أوجد قيمة مقدار ضغط الدم لإمرأة عمرها 46 سنة .

(iv) اوجد تقدير فترة الثقة لمعامل الانحدار β عند مستوى معنوية $\alpha = 0.05$.

(v) اختبر معنوية معامل الانحدار β عنـد مستـوى معنويـة $\alpha = 0.05$.

١٧ - الجدول التالي يمثل درجات الحرارة والمبيعات في إحدى محطات المحروقات

درجة الحرارة x	25	30	32	33	35	40	37
نمو النبات y	44	45	33	38	30	27	41

(i) أوجد معامل ارتباط بين درجة الحرارة والمبيعات بطريقتين مختلفتين.

(ii) أوجد معادلة خط انحدار y على x.

(iii) أوجد تقدير فترة الثقة لمعامل الانحدار β عند مستوى معنوية $\alpha = 0.05$.

الفصل الرابع عشر

الاختبارات اللامعلمية

Nonparametric Tests

● مقدمة ● اختبار الإشارة ● اختبار إشارة الرتب (ولكوكسون) ● اختبار مجموع الرتب (مان وتني) ● اختبار مجموع الرتب (كرسكال والس) ● اختبار العشوائية (الأشواط أو التكرارات المستمرة) ● تمارين

(١٤,١) مقدمة

تظهر في كثير من التطبيقات الإحصائية مواقف عديدة لا نعرف فيها توزيع المتغير محل الدراسة، ففي كثير من الأحيان نجد بيانات واقعية في البحوث الاجتماعية، أو اللغوية، أو الزراعية يصعب التعرف على الصيغة الدالية للتوزيع الاحتمالي التي تتبعه. في مثل هذه الحالات نلجأ عادة إلى ما يسمى بالاختبارات اللامعلمية (Nonparametric tests) وخاصة في العينات الصغيرة. حيث تتطلب بيانات العينات الصغيرة أن تكون مأخوذة من مجتمع طبيعي حتى يمكننا استخدام الاختبارات المعلمية السابقة (parametric tests) مثل اختبارات .....Z,t,F. كما تتطلب شروط إضافية بأن تكون العينات المأخوذة من المجتمع محل الدراسة مستقلة، وأحيانا تتطلب أن تكون تبايناتها متساوية. إن الاختبارات اللامعلمية تحل مكان الاختبارات المعلمية المعروفة حينما لا تستوفي تلك الاختبارات الشروط المطلوبة لتطبيقها. وسميت بالاختبارات اللامعلمية لأنها لا تختبر قيم المعالم لمجتمعات معروفة الشكل كما هو معروف في اختبارات ...Z,t,F.

يجب أن نلاحظ أن الاختبارات اللامعلمية أكثر عمومية من تلك التي تحتاج

إلى فروض إضافية، ولذا فمن المتوقع أنها لن تكون على نفس درجة كفاءة الاختبارات المعلمية خاصة في المشكلات التي يمكن أن تطبق فيها الاختبارات المعلمية. لذا يجب استخدامها فقط في الحالات التي يصعب فيها تطبيق الاختبارات المعلمية. سنستعرض في هذا الفصل بعض الاختبارات اللامعلمية التي تفيد الدارسين في كثير من المجالات محاولين شرح كيفية أداء الاختبار من خلال أمثلة توضح الاختبارات اللامعلمية وسنكتفي في هذا الفصل بدراسة بعض الاختبارات اللامعلمية مثل اختبار الإشارة واختبار إشارة الرتب (ولكوكسون) واختبار مجموع الرتب (مان هوتني) وتعميم اختبار مجموع الرتب لأكثر من مجتمعين وهو ما يسمى باختبار كرسكال والس ثم أخيرا اختبار العشوائية (الأشواط) وذلك بالشرح والتفصيل والأمثلة كما يلي:

(١٤,٢) اختبار الإشارة
Sign Test

في بعض الأحيان يصعب التحقق من أن المجتمع الذي اختيرت منه العينة يتبع توزيعا طبيعيا، لذلك فإن من الممكن استخدام اختبار الإشارة هو اختبار لامعلمي كبديل للاختبارات الخاصة بالمتوسط والاختبارات الخاصة بالفرق بين المتوسطات والمبنية على المشاهدات المزدوجة.

(١٤,٢,١) اختبار الإشارة في حالة عينة واحدة (Sign test for one sample)

يستخدم اختبار الإشارة في حالة عينة واحدة عندما تكون العينة مختارة من مجتمع متصل وحيث يكون احتمال الحصول على مشاهدة أقل من وسيط المجتمع يساوي احتمال الحصول على مشاهدة أكبر من وسيط المجتمع ولكل منهما احتمال $p = \frac{1}{2}$.

ولاختبار فرض العدم($H_0 : \varepsilon = \varepsilon_0$) ضد أي فرض بديل مناسب على

أساس عينة عشوائية حجمها n. فإننا نستبدل كل مشاهدة أكبر من ε_0 بإشارة موجبة وكل مشاهدة في العينة أقل من ε_0 بإشارة سالبة، (وإذا حدث أن وجدت مشاهدة تساوي ε_0 فإننا نهملها ولا نأخذها في الاعتبار).

ثم نختبر فرض العدم H_0 الذي ينص على أن الإشارات الموجبة مساوية للإشارات السالبة وهي قيم لمتغير عشوائي يتبع توزيع ذي الحدين بمعلمتين هما $n, p = \frac{1}{2}$. وعلى ذلك فإن الإحصاءة المناسبة هي X التي تمثل عدد الإشارات الموجبة لقيم متغير عشوائي يتبع توزيع ذي الحدين بمعلمتين $n, p = \frac{1}{2}$.

عند إجراء اختبار الإشارة لعينة واحدة حجمها صغير فإننا نستخدم توزيع ذي الحدين، وعندما يكون حجمها كبيرا نسبيا أي أن كل من $np, np(1-p)$ أكبر من 5 فإننا نستخدم التوزيع الطبيعي كتقريب لتوزيع ذي الحدين حيث المتغير X الذي يمثل عدد الإشارات الموجبة وله متوسط np وتباين $np(1-p)$ فإنه تحت صحة فرض العدم H_0 يمكن صياغة الإحصاءة Z والتي تتبع تقريبا توزيعا طبيعيا قياسيا وتحسب كما يلي:

$$Z = \frac{x - np}{\sqrt{np(1-p)}} = \frac{x - (1/2)n}{\sqrt{n(\frac{1}{2})(1-\frac{1}{2})}} = \frac{2x - n}{\sqrt{n}} \qquad (١٤,١)$$

مثال (١٤,١)

اختيرت عينة عشوائية صغيرة مكونة من 15 مشاهدة لمقدار الأكتين في أحد أنواع الجازولين وكانت المشاهدات كما يلي:

97.0, 94.7, 96.8, 95.5, 96.3, 99.8, 96.9, 94.8

92.7, 98.6, 95.6, 97.1, 97.7, 98.0, 94.4

والمطلـوب اختبـار ما إذا كان وسيط مقدار الأكتين يقل عـن 98.0 في الجازولين في وحدة القياس المستخدمة في الدراسة وذلك عند مستوى معنوية $\alpha = 0.01$.

الحل

أولا صياغة الفرض الإحصائي:

$$H_0 : \varepsilon_0 = 98.0$$

$$H_1 : \varepsilon_1 < 98.0$$

ثانيا: اختيار الإحصاءة ثم حسابها

باعتبار أن فرض العدم H_0 صحيح، فإننا نلاحظ أن هناك مشاهدة تساوي 98.0 فلا بد من إهمالها وعلى ذلك فإن حجم العينة المستخدم هي $n = 14$ ويبني الاختبار على أساس عدد الإشارات الموجبـة x فإذا كان X يرمز لمتغير عشوائـي لتوزيع ذي الحدين بمعلمتين $n = 14, p = \frac{1}{2}$ وبمستوى معنوية $\alpha=0.01$. فإننا نحسب الإحصاءة x باستبدال كل مشاهدة أكبر من 98.0 بإشارة موجبة (+) وكل مشاهدة أصغر من 98.0 بإشارة سالبة فإننا نحصل على ما يلي:

- - - - - + - - - + - - - -

وبذلك فإن الإحصـاءة $x = 2$ إشارة موجبة ويتم استخـدام x بأسلوبين لاختبـار الإشارة كما يلي:

(i) **الأسلوب الأول** وهو استخدام توزيع ذي الحدين وحساب الاحتمال $P(X < x)$ كما يلي:

$$P(X < x) = \sum_{X=0}^{2} \binom{14}{x} \left(\frac{1}{2}\right)^{x} \left(\frac{1}{2}\right)^{14-x}$$

$$= \binom{14}{0} \left(\frac{1}{2}\right)^{0} \left(\frac{1}{2}\right)^{14} + \binom{14}{1} \left(\frac{1}{2}\right)^{1} \left(\frac{1}{2}\right)^{13} + \binom{14}{2} \left(\frac{1}{2}\right) \left(\frac{1}{2}\right)^{13}$$

$$= 0.000061 + 0.000854 + 0.005554 = 0.006596 = 0.007$$

(ii) **الأسلوب الثاني** وهو استخدام التوزيع الطبيعي كتقريب لتوزيع ذي الحدين حيث إن كل من $np, np(1-p)$ تقريبا أكبر أو تساوي 5 فإننا نحسب الإحصاءة Z والمعطاة بالعلاقة (١ , ١٤) كما يلي:

$$Z = \frac{2x - n}{\sqrt{n}} = \frac{2(2) - 14}{\sqrt{14}} = \frac{-10}{3.74} = -2.67$$

ثالثا: تحديد منطقة قبول ورفض H_0.

(i) بالنسبة للأسلوب الأول فإن فترة قبول H_0 هي $(0.01, 1)$ وفترة رفض H_0 عندما $P(X \le x) \le 0.01$ أي الفترة $(0,0.01)$.

(ii) بالنسبة للأسلوب الثاني وهو استخدام الطبيعي كتقريب ذي الحدين والاختبار من طرف واحد فإن فترة قبول H_0 هي $(-2.33, \infty)$ ورفض H_0 خارج هذه الفترة.

رابعا: القرار الإحصائي

بالنسبة للأسلوب الأول (i) نجد أن قيمة الاحتمال $P(X \le 2) = 0.007$ واقعة في فترة رفض H_0.

وكذلك بالنسبة للأسلوب الثاني (ii) فإن الإحصاءة $Z = -2.67$ المحسوبة واقعة في منطقة رفض H_0 أي أن القرار في الأسلوبين (i) و (ii) هو رفض H_0 وقبول الفرض البديل H_1 أي أن مقدار وسيط الأكتين في الجازولين أقل من 98.0 في وحدة القياس المستخدمة.

(٢ , ٢ , ١٤) اختبار الإشارة في حالة العينة المزدوجة (Sign test for dauble sample)

لاختبار الإشارة أهمية كبيرة في المسائل التطبيقية ذات المشاهدات المزدوجة. على سبيل المثال إذا أردنا أن نعرف ما إذا كان نوع الغذاء A له تأثير أكبر في زيادة إنتاج اللبن لعينة من الأبقار عن إنتاج اللبن لنفس العينة من الأبقار باستخدام الغذاء B. حيث يتم وضع إنتاج كل بقرة من اللبن بواسطة الغذاء A أولا ثم بواسطة الغذاء B ثانيا في زوج واحد من المشاهدات لكل بقرة، ويتم ذلك لجميع الأبقار بالعينة. والمطلوب اختبار فرض العدم $H_0 : \varepsilon_1 = \varepsilon_2$ ضد الفرض البديل $H_1 : \varepsilon_1 > \varepsilon_2$.

فإنه يمكن استبدال كل زوج من المشاهدات بإشارة موجبة إذا كانت المشاهدة الأولى أكبر من المشاهدة الثانية وبإشارة سالبة إذا كانت المشاهدة الأولى أصغر من المشاهدة الثانية. ونهمل الزوج الذي فيه المشاهدتان متساويتان، ومن ثم نتبع الخطوات التي تم اتباعها في حالة العينة الواحدة في البند السابق ونوضح ذلك من خلال المثال التالي.

مثال (٢ , ١٤)

استخدم عقار جديد لعلاج مرض السكر، والمطلوب معرفة ما إذا كان له تأثير على ضربات القلب فأعطي العلاج لستة عشر مريضا بالسكر وكان عدد ضربات القلب في الدقيقة لكل منهم قبل وبعد العلاج هي:

(76,78), (80,81), (91,92), (75,74), (81,84), (77,77), (79,78), (82,83)

(88,83), (81,80), (78,79), (85,85), (76,74), (79,82), (82,79), (74,76)

استخدم مستوى معنوية $\alpha = 0.05$.

الحل

أولا: صياغة الفرض الإحصائي.

$$H_0: \varepsilon_1 = \varepsilon_2$$

$$H_1: \varepsilon_1 \neq \varepsilon_2$$

ثانيا: اختيار الإحصاءة ثم حسابها

باعتبار أن فرض العـدم H_0 صحيح، فإننا نلاحظ أن هناك زوجـين مـن المشاهدات هما (77,77) و (85,85) كل من القراءتين قبل وبعد العلاج متساويتين فلا بد من إهمالهما وعلى ذلك فإن حجم العينة المستخدمة في الاختبار هي $n =$ 14 ويبنى الاختبار على أساس x عدد الإشارات الموجبة حيث المتغير العشوائي X له توزيع ذي الحدين $P = \frac{1}{2}$ و $n=14$ ثم نحسب الإحصـاءة x باستبدال كل زوج بإشارة موجبة عندما تكون قراءته الأولى أكبر من الثانية وبإشارة سالبة عـنـدمـا قراءته الأولى أصغر من الثانية، وبذلك نحصل على الإشارات لمثالنا كما يلي:

-, -, -, +, -, 0, +, -, +, +, -, 0, +, -, +, -

وبذلك فإن الإحصـاءة $x = 6$ إشارة موجبة ويتم استخـدام x بأسلوبين لاختبـار الإشارة كما يلي:

(i) الأسلوب الأول وهو استخدام توزيع ذي الحدين وحساب الاحتمال $P(X \leq x)$ كما يلي:

$$P(X \leq x) = \sum_{x=0}^{6} \binom{14}{x} \left(\frac{1}{2}\right)^{x} \left(\frac{1}{2}\right)^{14-x}$$

من جدول توزيع ذي الحدين (٨) آخر الكتاب نجد أن

$$P(X \leq x) = 0.000 + 001 + 0.006 + 0.022 + 0.122 + 0.183 = 0.334$$

(ii) الأسلوب الثاني وهو استخدام التوزيع الطبيعي لتقريب ذي الحدين لأن كل من np, np(1-p) تقريبا يساوي أو أكبر من 5، ويتم حساب الإحصاءة Z المعطاة بالعلاقة (١ , ١٤) السابقة كما يلي:

$$Z = \frac{2x - n}{\sqrt{n}} = \frac{2(6) - 14}{\sqrt{14}} = \frac{-2}{3.74} = 0.53$$

ثالثا: تحديد منطقة قبول ورفض H_0.

(i) بالنسبة للأسلوب الأول فإن فترة قبول H_0 هي (0.05, 0.95) أي أن $0.05 \le P(X \le 6) \le 0.95$ ومنطقة رفض H_0 خارج هذه الفترة.

(ii) بالنسبة للأسلوب الثاني فإن قبول H_0 عندما Z المحسوبة تقع داخل الفترة (-1.96, 1.96) ورفض H_0 خارج هذه الفترة.

رابعا: القرار الإحصائي.

بالنسبة للأسلوب الأول (i) نجد أن قيمة الاحتمال $P(X \le 6) = 0.334$ واقعة في فترة قبول H_0.

وكذلك بالنسبة للأسلوب الثاني (ii) فإن الإحصاءة $Z = 0.53$ واقعة داخل فترة قبول H_0 أي أن القرار في كلا الأسلوبين (i) و (ii) واحد وهو قبول H_0 أي أنه لا يوجد تأثير للعقار على زيادة ضربات القلب على المرضى عند استخدامه وذلك بمستوى معنوية $\alpha = 0.05$.

(١٤,٣) اختبار إشارة الرتب (ولكوكسون)

Rank Sign Test (Wilcoxon)

لاحظنا أن اختبار الإشارة كان يعتمد على عدد الإشارات الموجبة للفرق بين مشاهدات العينة والمتوسط ε_0 في حالة العينة الواحدة أو عدد الإشارات الموجبة للفرق بين أزواج المشاهدات في حالة العينة المزدوجة ولا يأخذ في الاعتبار مقدار هذا الفرق. ولكن اختبار إشارة الرتب يأخذ في الاعتبار نوع الإشارة ومقدار هذا الفرق. ويرجع هذا الاختبار إلى العالم ولكوكسون (Wilcoxon) ويستخدم اختبار إشارة الرتب في حالة عينة واحدة عندما تكون العينة مختارة من مجتمع متصل ومتماثل وتحت صحة فرض العدم $H_0 : \varepsilon = \varepsilon_0$ فإننا نقوم بطرح ε_0 من كل مشاهدة من المشاهدات مع إهمال الفروق التي تساوي أصفارا. ثم نقوم بترتيب هذه الفروق الباقية بدون النظر إلى الإشارة، فنعطي الرتبة 1 لأقل فرق مطلق (بدون النظر إلى إشارة الفرق) والرتبة 2 للفرق الذي يليه وهكذا لباقي الفروق، وفي حالة ما إذا كان هناك فروق متساوية (بدون النظر إلى الإشارة) فإننا نعطيها متوسط الرتبتين والتي كانت ستأخذها لو أنها كانت مختلفة (ذلك مثل ما تم حسابه في حالة ارتباط الرتب في الفصل الثالث عشر السابق). وتحت صحة فرض العدم H_0 فإن مجموع الرتب المناظرة للإشارات الموجبة يجب أن يساوي تقريبا مجموع الرتب المناظرة للإشارات السالبة. ضد أي فرض بديل مناسب $H_1(\varepsilon \neq \varepsilon_0;\ \varepsilon > \varepsilon_0;\ \varepsilon < \varepsilon_0)$ ونختار الإحصاءة T = مجموع رتب الإشارات الموجبة وإذا كان حجم العينة $n \geq 15$ فإن الإحصاءة T تقترب من التوزيع الطبيعي بمتوسط $\mu_T = \dfrac{n(n+1)}{4}$ وتباين $\sigma_T^2 = \dfrac{n(n+1)(2n+1)}{24}$ ثم نستخدم العلاقة (١٤,٢) التالية للإحصاءة Z لاختبار إشارة الرتب:

$$Z = \frac{T - \mu_T}{\sigma_T} \qquad (١٤,٢)$$

وفي حالة استخدام العينة المزدوجة فإن فرض العدم H_0 يصبح $H_0 : \varepsilon_1 = \varepsilon_2$ حيث نقوم بطرح المشاهدة الثانية من المشاهدة الأولى لكل أزواج المشاهدات في العينة المزدوجة مع اهمال الأزواج الذين تتساوى فيهم قيم المشاهدة الأولى والثانية لكل زوج. ثم نرتب الفروق ونتبع نفس الخطوات السابقة مثل ما تم في حالة العينة الواحدة وباستخدام العلاقة (١٤,٢) للإحصاءة Z لاختبار إشارة الرتب ونوضح ذلك بالأمثلة كما يلي.

مثال (١٤,٣)

استخدم بيانات مثال (١٤,١) السابق بتواجد الأكتين لأحد أنواع الجازولين بوسيط يقل عن 98.0 عند مستوى معنوية $\alpha = 0.01$. باستخدام اختبار إشارة الرتب (ولكوكسون).

الحل

أولا: صياغة الفرض الإحصائي.

$$H_0 : \varepsilon = 98.0$$

$$H_1 : \varepsilon < 98.0$$

ثانيا: اختيار الإحصاءة ثم حسابها

حيث إن عدد البيانات للعينة كبيرا $n = 15$ فإنه يمكن اختيار الإحصاءة T = مجموع الرتب الموجبة، والتي تقترب من التوزيع الطبيعي بمتوسط وتباين هما $\mu_T = \frac{n(n+1)}{4}$، $\sigma_T^2 = \frac{n(n+1)(2n+1)}{24}$ ويمكن استخدام الإحصاءة Z المعطاة بالعلاقة (١٤,٢) لإشارة الرتب ويتم حساب T، μ_T، σ_T و Z كما هو موضح في جدول (١٤,١) كما يلي:

جدول (١ , ١٤)

قيم الأكتين في العينة	فروق القيم عند $\mu_0 = 98.0$	الرتب	
97.0	-1	4	$n = 14$
94.7	-3.3	12	مجموع الرتب الموجبة = T
96.8	-1.2	6	$T = 8+2 = 10$
95.5	-2.5	10	$\mu_T = \frac{n(n+1)}{4} = \frac{14(15)}{4} = 52.5$
96.3	-1.7	7	
99.8	+1.8	8	$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$
96.9	-1.1	5	
94.8	-3.2	11	$= \sqrt{\frac{14(15)(29)}{24}} = 15.93$
92.7	-5.3	14	واستخدام العلاقة (٢ , ١٤)
98.6	+0.6	2	فإن
95.6	-2.4	9	$Z = \frac{T - \mu_T}{\sigma_T} = \frac{10 - 52.5}{15.93}$
97.1	-0.9	3	
97.7	-0.3	1	$= -2.668$
98.0	0	--	
94.0	-5.0	13	

ثالثا: تحديد منطقة رفض وقبول H_0 .

نحدد فترة قبول H_0 حيث الفرض البديل H_1 من طرف واحد وهو الأيسر . فإن فترة قبول H_0 عند مستوى معنوية $\alpha = 0.01$ هي $(-2.33, \infty)$ ومنطقة رفض H_0 هي خارج هذه الفترة.

رابعا: القرار الإحصائي

حيث إن قيمة $Z = -2.668$ المحسوبة واقعة خارج فترة قبول H_0، فإننا نرفض H_0 أي أن متوسط الأكتين في الجازولين أقل من 98.0 وذلك بدرجة ثقة 99%.

مثال (٤ , ١٤)

استخدم بيانات مثال (٢ , ١٤) العقار عند مستوى معنوية $\alpha = 0.05$ وذلك باستخدام اختبار إشارة الرتب.

الحل

أولا: صياغة الفرض الإحصائي

$H_0 : \varepsilon_1$ (وسيط ضربات القلب قبل العقار) = ε_2 (وسيط الضربات بعد العقار)

$$H_1 : \varepsilon_1 \neq \varepsilon_2$$

ثانيا: اختيار الإحصاءة ثم حسابها

حيث إن عدد بيانات العينة كبيرا وهو $n = 16$، فإنه يمكن استخدام الإحصاءة T = مجموع الرتب الموجبة، والتي تقترب من التوزيع الطبيعي بمتوسط $\mu_T = \frac{n(n+1)}{4}$ وتباين $\sigma_T^2 = \frac{n(n+1)(2n+1)}{24}$ وبالتالي يمكن استخدام الإحصاءة Z المعطاة بالعلاقة (٢ , ١٤) لاختبار إشارة الرتب كما هو موضح في جدول (٢ , ١٤) كما يلي:

جدول (٢ , ١٤)

ضربات القلب		فرق المشاهدات	الرتب	
قبل العقار	بعد العقار	قبل - بعد		
76	78	-2	9	$n = 14$
80	81	-1	4	مجموع الرتب الموجبة = T
91	92	-1	4	$T = 4+4+14+4+12+9$
75	74	+1	4	$= 47$
81	84	-3	12	$\mu_T = \frac{n(n+1)}{4} = \frac{14(15)}{14}$
77	77	--	--	$= 52.5$
79	78	+1	4	
82	83	-1	4	$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$
88	83	+4	14	
81	80	+1	4	$= \sqrt{\frac{14(15)(29)}{24}}$
78	79	-1	4	
85	85	--	--	$= 15.93$
76	74	+2	9	ونحسب من العلاقة (٢ , ١٤)
79	82	-3	12	فإن
82	79	+3	12	$Z = \frac{47 - 52.5}{15.93}$
74	76	-2	9	$= - 0.345$

ثالثا: تحديد منطقة رفض وقبول H_0.

نحدد فترة قبـول H_0 حيث إن الفرض البديـل H_1 من طرفين. فإن فترة قبول H_0 هي $(-1.96, +1.96)$ عند مستوى معنوية $\alpha = 0.05$ ورفض H_0 خارج هذه الفترة.

رابعا: القرار الإحصائي

حيث إن $Z=-0.345$ المحسوبة واقعة داخل فترة قبول H_0 أي أن لا تأثير للعقار على ضربات القلب بدرجة ثقة 95%.

(٤ , ١٤) اختبار مجموع الرتب (اختبار مان وتني)

Sum of Rank Test Mann-Whitney Test

يعتبر اختبار مان وتني الاختبار اللا معلمي الخاص بالفرق بين وسطي أو متوسطي مجتمعين والمبني على أساس عينتين مستقلتين، أي أن هذا الاختبار بديل لاختبار t بل أنه أفضل منه خاصة إذا كانت العينتين مختارتين من مجتمعين لا يتبعان توزيعا طبيعيا ويرجع هذا الاختبار إلى العالم مان وتني (Mann-Whitney).

ولتوضيح هذا الاختبار نفترض أن لدينا عينتين مستقلتين الأولى حجمها n_1 والثانية حجمها n_2 تم اختيارهما من مجتمعين متصلين الأول وسيطة ε_1 والثاني وسيطة ε_2 والمطلوب اختبار فرض العدم $H_0 : \varepsilon_1 = \varepsilon_2$ ضد الفرض البديل H_1 $(\varepsilon_1 \neq \varepsilon_2$ أو $\varepsilon_1 > \varepsilon_2$ أو $\varepsilon_1 < \varepsilon_2)$.

وفي هذا الاختبار نقوم بدمج المشاهدات للعينتين معا في عينة واحدة ثم نقوم بإعطاء رتب للمشاهدات تصاعديا فتعطي الرتبة 1 لأصغر مشاهدة في العينتين بعد الدمج والرتبة 2 التي تليها وهكذا حتى تعطى أكبر رتبة لأكبر مشاهدة وتحت صحة فرض العدم H_0 فإن متوسط الرتب لكل من العينتين يكون هو نفسه أي أن الرتب الكبرى والصغرى موزعة بالتساوي في كل من العينتين. أما تحت صحة الفرض البديل H_1 وهو عدم تساوي وسيطى المجتمعين $\varepsilon_1 \neq \varepsilon_2$ فإننا نجد أن معظم الرتب الصغيرة تذهب إلى أحد العينتين بينما معظم الرتب الكبيرة تذهب إلى العينة الأخرى.

واختبار مان وتني تحت صحة فرض العـدم H_0 يختار الإحصـاءة U التي تعطى بالعلاقة (١٤,٣) التـالـيـة:

$$U = n_1 n_2 + \frac{n_2(n_1 + 1)}{2} - W \qquad (١٤,٣)$$

حيث W = مجموع الرتب للعينة الأولى التي حجمها n_1.

وتحت صحة فرض العدم H_0 فإن متوسط الإحصاءة U وهو μ_u والانحراف المعياري لها وهـو σ_u حيث قيمتيهما كما يـلـي:

$$\mu_u = n_1 n_2 / 2 \qquad (١٤,٤)$$

$$\sigma_u = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \qquad (١٤,٥)$$

وعندمـا يكون حجم كل من العينتين كبيرا كبرا كافيا بحيث يكون حجم كل منهما أكبر من 8 مشاهدات فإن الإحصاءة U تقترب من التوزيع الطبيعي وبذلك يمكننا استخدام الإحصـاءة Z المعطاة بالعلاقة (١٤,٦) التالـيـة:

$$Z = \frac{U - \mu_u}{\sigma_u} \qquad (١٤,٦)$$

ونوضح طريقة استخدام اختبار مان وتني بالمثال التالي.

مثال (١٤,٥)

استخدم اختبار مان وتني عند مستوى معنوية $\alpha = 0.05$ لاختبار ما إذا كان وسيط أقطار حبات الرمال من موقعين مختلفين (A, B) على القمر يكونا مختلفين حيث وجد أن البيانات لكل من الموقعين لأقطار حبات الرمال بالمليمتر كما يلي:

الموقــع A: 0.30, 0.43, 1.34, 0.18, 0.41, 0.47, 0.15, 0.38

0.82, 0.49, 0.45, 0.10, 0.16, 0.33, 0.61

الموقــع B: 1.09, 0.87, 0.99, 0.90, 0.37, 0.94, 1.11, 0.56

1.05, 0.46, 0.51, 0.81, 0.24, 0.52

الحل

أولا: صياغة الفرض الإحصائي.

$$H_0: \ \varepsilon_1 = \varepsilon_2$$
$$H_0: \ \varepsilon_1 \neq \varepsilon_2$$
$$\alpha = 0.01$$

ثانيا: اختيار الإحصاءة ثم حسابها.

لاستخدام اختبار مان وتني فإنه يلزم ايجاد الإحصاءة W وهي تمثل مجمــوع رتب العينة الأولــى$n_1 = 15$ بعد دمجها في العينة الثانيــة$n_2 = 14$ واعطاء رتب للمشاهدات بعد الدمج ثم حساب الإحصاءة U ومتوسطهاμ_u وانحرافها المعياريσ_u كما هو موضح بالعلاقات (٣, ١٤)، (٤, ١٤) و (٥, ١٤) ثم حساب الإحصاءة Z من العلاقة (٦, ١٤) للتوزيع الطبيعي ويتم ذلك كما في جدول (٣, ١٤) كما يلي.

ثالثا: تحديد منطقة قبول ورفض H_0.

نحدد فترة قبول H_0 بالفترة ($Z_{0.005}$ + و $Z_{0.005}$ -) وهي (1.96-, 1.96+) ومنطقة رفض H_0 خارج هذه الفترة.

جدول (٣, ١٤)

	رتـــب المشاهدات	المشاهدات المدمجة للعينتين
$n_1 = 15\ ,\ n_2 = 14$	6	0.30
مجموع الرتب للعينة الأولى = W		
W = 6+11+29+4+10+14+2+9+20	11	0.43
+15+12+1+3+7+19 = 162	29	1.34
	4	0.18
$U = n_1 n_2 + \frac{n_2(n_1 + 1)}{2} - W$	10	0.41
	14	0.47
	2	0.15
$= 15(14) + \frac{14(15 + 1)}{2} - 162$	9	0.38
	20	0.82
$= 168$	15	0.49
	12	0.45
$\mu_u = \frac{n_1 n_2}{2} = \frac{15(14)}{2} = 105$	1	0.10
	3	0.16
$\sigma_u = \sqrt{\frac{15(14)\ (15+14+1)}{12}} = 22.9$	7	0.33
ونحسب Z من العلاقة (٦, ١٤) فإن	19	0.61
	*	*
$Z = \frac{U - \mu_u}{\sigma_u} = \frac{160 - 105}{22.9}$	27	1.09
	22	0.87
	25	0.99
	23	0.90
$= 2.4017$	8	0.37
	24	0.94
	28	1.11
	18	0.56
	26	1.05
	13	0.46
	16	0.51
	21	0.86
	5	0.24
	17	0.52

رابعا: القرار الإحصائي

حيث إن قيمة $Z = 2.4017$ المحسوبة واقعة خارج منطقة قبول H_0 $(-1.96, 1.96)$ فإننا نرفض H_0 ونقبل الفرض البديل H_1 القائل بأن متوسطي قطري حبات الرمال في المنطقتين A, B مختلفين بدرجة ثقة 95%.

(١٤,٥) اختبار مجموع الرتب (كرسكال والس)

Rank Sum Test Kruskal-Wallis

يعتبر اختبار كرسكال والس هو تعميم لاختبار مان وتني لأكثر من عينتين مستقلتين تحت صحة فرض العدم H_0، وعندما يكون لدينا k عينة $(k \geq 2)$ مأخوذة من مجتمعات لها نفس المتوسط أو الوسيط ضد الفرض البديل H_1 القائل بأن المتوسطات لهذه المجتمعات ليست كلها متساوية. وهذا الاختبار هو بديل عن اختبار F لتحليل التباين في اتجاه واحد والذي يتطلب أن المجتمعات تكون تقريبا لها توزيع معتدل ولها نفس الانحراف المعياري.

في اختبار كرسكال والس لا يتطلب شروط اختبار F السابقة ويتم في اختبار كرسكال والس دمج العينات المستقلة في عينة واحدة حجمها n حيث $n = n_1 + n_2 + ... + n_k$ ويفضل بأن يكون حجم كل عينة 5 مشاهدات فأكثر. ثم نقوم باعطاء رتب للمشاهدات تصاعديا فتعطى الرتبة 1 لأصغر مشاهدة في العينات بعد الدمج والرتبة 2 للمشاهدة التي تليها وهكذا لباقي المشاهدات. ثم تحسب القيمة W_1 = مجموع رتب العينة الأولى، W_2 = مجموع رتب العينة الثانية وهكذا W_k = مجموع رتب العينة k. وذلك مثل ما تم في اختبار مان وتني بند (١٤,٤) السابق ثم تحسب الإحصاءة بالعلاقة (١٤,٧) التالية:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{W_i^2}{n_i} - 3(n+1) \qquad (١٤,٧)$$

وتحت صحة فرض العدم H_0 فإن الإحصاءة H المعطاة بالعلاقة (١٤,٧) تقترب

من توزيع مربع كاي بدرجات حرية k - 1 ونرفض فرض العدم $H_0 : \varepsilon_1 = \varepsilon_2 = \cdots = \varepsilon_k$ عند مستوى معنوية α عندما تقع قيمة H المحسوبة بالعلاقة (١٤,٧) خارج فترة قبول H_0 وهي $(0, \chi^2_{\alpha, k-1})$ ونوضح ذلك بالمثال التالي.

مثال (١٤,٦)

نفترض أنه لدينا ثلاثة مجموعات من الدارسين للغة العربية من غير الناطقين بها واستخدم في تعليمهم للغة العربية ثلاث طرق مختلفة خلال الفصل الدراسي ومعامل اللغات وكان الاختبار موحد للطرق الثلاث المختلفة وكانت نتائج الاختبار بالدرجات كما يلي:

الطريقة الأولى A :	94, 87, 91, 47, 87, 97
الطريقة الثانية B :	85, 82, 79, 84, 61, 72, 80
الطريقة الثالثة C :	89, 67, 72, 76, 69

والمطلوب اختبار ما إذا كانت هذه الطرق مختلفة في التعليم عند مستوى معنوية $\alpha = 0.05$.

الحل

أولا: صياغة الفرض الإحصائي.

$$H_0 : \varepsilon_1 = \varepsilon_2 = \varepsilon_3$$

$H_1 : \varepsilon_i$ على الأقل إحدى الطرق مختلفة في متوسط الدرجات

ثانيا: اختيار الإحصاءة ثم حسابها.

والمطلوب حساب الإحصاءة H المعطاة بالعلاقة (٧ , ١٤) السابقة، ولحسابها ندمج بيانات العينات الثلاثة السابقة ثم نوجد لها الرتب مثل ما تم في اختبار مان وتني السابق ولتوضيح ذلك من خلال جدول (٤ , ١٤) كما يلي:

جدول (٤ , ١٤)

	رتب المشاهدة	قيم المشاهدة
$n_1 = 6, W_1 = 17+13.5+16+6+13.5+18$	17	94
$= 84$	13.5	87
$n_2 = 7, W_2 = 12+11+8+10+1+4.5+9$	16	91
$= 55.5$	6	74
$n_3 = 5, W_3 = 15+2+4.5+7+3$	13.5	87
$= 31.5$	18	97
$n = n_1+n_2+n_3 = 6+7+5 = 18$		*
ثم نحسب الاحصاءة H من العلاقة (٧ , ١٤) كما يلي :	12	85
$H = \frac{12}{n(n+1)} \sum_{i=1}^{3} \frac{W_i}{n_i} - 3(n+1)$	11	82
	10	84
$= \frac{12}{18(19)} (\frac{(84)^2}{6} + \frac{(55.5)^2}{7} + \frac{(31.5)^2}{5} - 3(19))$	1	61
$= \frac{2}{342} (1176+436.1+198.5) - 57$	4.5	72
	9	80
$= 63.57 - 57$		**
$= 6.53$	15	89
	2	67
	4.5	72
	7	76
	3	69

ثالثا: تحديد منطقة رفض وقبول H_0.

تحدد منطقة قبول H_0 بالفترة$(0, \chi^2_{0.05}(2))$ ومن جدول (٣) لمربع كاي آخر الكتاب والتي تصبح الفترة التالية $(0, 5.991)$ ومنطقة رفض H_0 هي خارج هذه الفترة عند مستوى معنوية $\alpha = 0.05$.

رابعا: القرار الإحصائي.

حيث إن الإحصاءة $H_0 = 6.53$ المحسوبة واقعة خارج فترة قبـول H_0 والتي هي $(0, 5.991)$ فإننا نرفض H_0 ونقبل الفرض البديل القائل أن طرق التعليم عـلـى الأقل أحدها مختلف.

(٦ , ١٤) اختبار العشوائية (الأشواط أو التكرارات المستمرة)

Tests of Randomness (Runs)

تعتبر كل الطرق التي تم مناقشتها فيما سبق أن العينات مأخوذة بـطـريـقـة عشوائية ولكن توجد طرق عملية يكون من الصعب الحكم على أن العينة المأخوذة هي عشوائية أم لا. واختبار العشوائية (االأشواط) يقيس العشوائية لبيانات العينة بعد اخذها من المجتمع محل الدراسة ونوضح اختبار العشوائية بالمثال التالي:

مثال (٧ , ١٤)

اعتبر الترتيب التالي للأشجار السليمة H والأشجار المريضة D بإحدى الطرق والمزروعة لفترة طويلة بأحد الأقاليم وكان الترتيب كما يلي:

<u>HHHHH</u> <u>DDD</u> <u>HHHHHHHHH</u> <u>DD</u> <u>HH</u> <u>DDDDD</u> <u>HHHH</u> .

نلاحظ من البيانات السابقة والتي تحتها خط أن أول التكرارات المستمرة تتألف من خمسة H والثانية من ثلاثة D والثالثة من تسعة H والرابعة من إثنين D والخامسة من أثنين H والسادسة من خمسة D والسابعة من أربعة H.

عدد الأشواط في المثال السابق يشتمل على مجموع سبعة تكرارات مستمرة ومجموع التكرارات المستمرة يستخدم كمؤشر لقياس العشوائية للبيانات. فمثلا إذا كان مجموع الأشواط قليل فيعتبر كمؤشر على وجود مجموعات عنقودية أو ربما اتجاه معين. وعندما يكون مجموع التكرارات المستمرة كثيرا. ربما يجعلنا نعتقد أنه يوجد نوع من الدورية. وفي المثال السابق تظهر الأشجار المريضة أنها توجد في مجموعات عنقودية والآن مطلوب اختبار أن هذه المجموعات العنقودية معنوية أم راجعة إلى الصدفة والعشوائية.

والاختبار الذي سوف يستخدم يعتمد على تراتيب n_1 حرف من أحد الأنواع مع n_2 حرف من النوع الآخر ويستخدم الإحصاءة U = مجموع التكرارات المستمرة في العينة $n = n_1 + n_2$ حيث يكون لها متوسط μ_u وانحراف معياري σ_u كما يلي:

$$\mu_u = \frac{2n_1 n_2}{n_1 + n_2} + 1 \qquad (١٤,٨)$$

$$\sigma_u = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 + 1)}} \qquad (١٤,٩)$$

وعندما يكون كل من n_1 و n_2 كبيرة على الأقل كل منهما لا يقل عن عشرة فإن التوزيع العيني للإحصاءة U يقترب من التوزيع الطبيعي وتحسب الإحصاءة Z كما يلي:

$$Z = \frac{u - \mu_u}{\sigma_u} \qquad (١٤,١٠)$$

وأن منطقة قبول H_0 (العينة عشوائية: H_0) هي الفترة $(-Z_{\alpha/2}, +Z_{\alpha/2})$ عند مستوى معنوية α ومنطقة رفض H_0 خارج هذه الفترة.

مثال (١٤,٨)

اختبر مثال (١٤,٧) للأشجار السليمة H والاشجار المريضة D ما إذا كان فرض العدم H_0 بأن الأشجار المريضة موجودة بطريقة عشوائية عند مستوى معنوية $\alpha = 0.01$.

الحل

أولا: صياغة الفرض الإحصائي.

الأشجار المريضة موجودة بطريقة عشوائية : H_0

الأشجار المريضة موجودة بطريقة غير عشوائية: H_1

ثانيا: اختيار الإحصاءة ثم حسابها.

نحسب n_1 وهي تمثل عدد الأشجار السليمة H أي أن $n_1 = 20$

ونحسب n_2 وهي تمثل عدد الأشجار المريضة D أي أن $n_2 = 10$

ونحسب مجموع التكرارات المستمرة u أي أن $u = 7$

ثم نحسب المتوسط μ_u والانحراف المعيـاري σ_u باستخدام العلاقتين (١٤,٨) و (١٤,٩) كما يلي:

$$\mu_u = \frac{2n_1 n_2}{n_1 + n_2} + 1 = \frac{2(20)(10)}{20+10} + 1 = 14.33$$

$$\sigma_u = \sqrt{\frac{1(20)(10)(2(20)(10) - 20 - 10)}{(20 + 10)^2 (20 + 10 - 1)}} = 2.38$$

ونحسب الإحصاءة Z من العلاقة (١٤,١٠) كما يلي:

$$Z = \frac{u - \mu_u}{\sigma_u} = \frac{7 - 14.33}{2.38} = -3.08$$

ثالثا: تحديد منطقة قبول ورفض H_0.

نحدد فترة قبول H_0 بالفترة $(-Z_{\alpha/2}, Z_{\alpha/2})$ وحيث إن مستوى المعنوية $\alpha = 0.01$ فإن فترة قبول H_0 هي $(-2.56, 2.56)$ ومنطقة رفض H_0 هي خارج هذه الفترة.

رابعا: القرار الإحصائي.

وحيث إن قيمة $Z_0 = -3.08$ المحسوبة واقعة خارج فترة قبول H_0 والتي هي $(-2.56, 2.56)$ أي أننا نرفض H_0 القائل أن الاشجار المريضة موجودة بطريقة عشوائية وذلك عند مستوى معنوية $\alpha = 0.01$.

ملاحظة

يستخدم اختبار العشوائية (الأشواط) في حالة البيانات العددية لاختبار ما إذا كانت مشاهدات العينة موجودة بصورة عشوائية أم لا ويتم ذلك بحساب الوسيط للعينة ثم إعطاء المشاهدة التي قيمتها أكبر من الوسيط بالرمز a والتي أصغر من الوسيط بالرمز b وبذلك تتحول بيانات العينة إلى رموز من a أو b ثم نحسب عدد الرموز a وليكن n_1 وعدد الرموز b وليكن n_2 والعينة كلها $n = n_1 + n_2$ ونحسب عدد التكرارات المستمرة u ثم μ_u و σ_u ثم الإحصاءة Z من العلاقات (١٤ , ٨)، (١٤ , ٩) و (١٤ , ١٠) السابقة ونوضح ذلك بالمثال التالي.

مثال (١٤ , ٩)

إذا أخذنا 40 رحلة متتالية للحافلات بين مدينتين وكان عدد الركاب لكل حافلة في كل رحلة هو كما يلي:

21,19,24,33,23,22,26,29,31,13,28,24,21,10,28,18,19,27,28,30,23,27

32,23,20,26,28,21,24,17,23,14,27,22,35,17,24,21,28,18

والمطلوب اختبار ما إذا كان يمكن اعتبار بيانات هذه العينة بيانات عينة عشوائية عند مستوى معنوية $\alpha = 0.05$ حيث وسيط هذه البيانات هو 23.5.

الحل

أولا: صياغة الفرض الإحصائي.

H_0 : بيانات العينة عشوائية

H_1 : بيانات العينة غير عشوائية

ثانيا: اختيار الإحصاءة ثم حسابها

نحسب عدد الأشواط u والتي تتم بأن نضع الرمز a للمشاهدات التي قيمتها أكبر من الوسيط 23.5 ونضع الرمز b للمشاهدات التي قيمتها أقل من الوسيط 23.5 فنحصل على الرموز التالية لمشاهدات العينة:

bbaabbaaabaabbabbaaabaabbaababbbabababab

فإنه يكون لدينا عدد الرمز a هو $20 = n_1$

وعدد الرمز b هو $20 = n_2$

ومجموع الأشواط u = 25 مرة

ونحسب μ_u و σ_u من العلاقتين (٨, ١٤) و (٩, ١٤) كما يلي:

$$\mu_u = \frac{2n_1 n_2}{n_1 + n_2} + 1$$

$$\mu_u = \frac{1(20)(20)}{20+20} + 1 = 21$$

$$\sigma_u = \sqrt{\frac{2n_1 n_2(2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 + 1)}}$$

$$= \sqrt{\frac{2(20)(20)(2(20)(20) - 20 - 20)}{(20 + 20)^2 (20+20+1)}} = 3.12$$

ثم نحسب الإحصاءة Z من العلاقة (١٠ , ١٤) كما يلي:

$$Z = \frac{u - \mu_u}{\sigma_u} = \frac{25 - 21}{3.12} = 1.28$$

ثالثا: تحديد منطقة قبول ورفض H_0

نحدد منطقة قبول H_0 بالفترة $(-Z_{\alpha/2}, Z_{\alpha/2})$ وحيث $\alpha = 0.05$ فإن فترة قبول H_0 هي $(-1.96, 1.96)$ ومنطقة رفض H_0 خارج هذه الفترة.

رابعا: القرار الإحصائي

حيث إن قيمة $Z_0 = 1.28$ المحسوبة واقعة داخل فترة قبول H_0، فإننا لا نستطيع رفض H_0 أي أن العينة عشوائية عند مستوى معنوية $\alpha = 0.05$.

(١٤ , ٧) تمارين

١ - في دراسة وقت انتظار الحافلة للعمال التي يستخدمونها في الذهاب إلى العمل في إحدى المؤسسات سجلت فترة للحافلة بالدقائق لعينة من 15 يوما فوجد أن زمن الانتظار بالدقائق كما يلي:

4,8,7,7,2,6,8,5,9,6,1,5,6,5,9

اختبر الفرض القائل بأن وسيط انتظار العامل للحافلـة 5 دقائق وذلك عند مستوى معنويـة $\alpha = 0.05$ وذلك باستخدام اختبار الإشارة ثم اختبار إشـارة الرتب.

٢ - في تجربة معملية لدراسة معامل الاحتكاك بين أحد المعادن والجلد اخـذت النتائج التالية:

0.59, 0.48, 0.53, 0.57, 0.55, 0.61, 0.56, 0.60, 0.55
0.58, 0.52, 0.55, 0.56, 0.51, 0.56, 0.58, 0.59

استخدم اختبار الإشارة عند مستوى معنوية $\alpha = 0.05$ لاختبار فرض العدم H_0 بأن وسيط معامل الاحتكاك $\varepsilon_0 = 0.55$ ضد الفرض البديل H_1 بأن وسيط معامل الاحتكاك $\varepsilon \neq 0.55$ ثم استخدم أيضا اختبار إشارة الرتب (ولكوكسون).

٣ - في دراسة لمعرفة تأثير برنامج للأمن الصناعي في عينة مكونة من 12 مصنعا سجلت عدد ساعات العمل المهدرة خلال أسبوع بسبب الحوادث داخل هذه المصانع قبل وبعد تنفيذ البرنامج وكانت النتائج كما يلي:

(37,28), (72,59), (26,24), (125,120), (45,46), (54,43), (13,15)
(79,75), (12,18), (34,29), (39,35), (26,24)

استخدم اختبار الإشارة مرة واختبار إشارة الرتب مرة أخرى عند مستـوى معنوية $\alpha = 0.05$، لاختبار ما إذا كان برنامج الأمن الصناعي له تأثير معنوي في تقليل وسيط الساعات المهدرة أسبوعيا.

٤ - البيانات التالية أخذت لمعرفة عما إذا كان هناك اختلاف معنوي في الوزن بين نوعين من الموازين لعينات من الصخور تم وزنها بالميزانين وكانت النتائـج بالجرام كما يلي:

العينة	الوزن بالميزان الأول	الوزن بالميزان الثاني
العينة الأولى	12.13	12.17
" الثانية	17.56	17.16
" الثالثة	11.40	11.42
" الرابعة	28.62	28.61
" الخامسة	10.25	10.27
" السادسة	23.37	23.42
" السابعة	16.27	26.26
" الثامنة	12.40	12.45
" التاسعة	24.78	24.75
" العاشرة	9.33	9.35

استخدم اختبار الإشارة عند مستوى معنويـة $\alpha = 0.01$ لاختبار عما إذا كان هناك فرق معنوي بين وسيطي الوزن.

٥ - في دراسة مرتبطة بجداول الحياة للأعمار التالية سجلت لعينتين عشوائيتين:

العينة الأولى: 78,54,67,69,57,75,92,71,77,79,73,83,60,85,55

العينة الثانية: 65,84,76,82,72,81,53,89,56,70,66,74

استخدم اختبار مان وتني عند مستوى معنوية $\alpha = 0.05$ لاختبار عما إذا كانت العينتين من مجتمعين مختلفين.

٦ - البيانات التالية تمثل الأعمار بالساعات لنوعين من المصابيح المستخدمة بصفة مستمرة

النوع الأول: 407,426,453,378,434,396,
441,373,393, 386,415, 418

النوع الثاني: 403,424,383,445,439,417,
412,462,439,432,413,433

استخدم اختبار مان وتني عند مستوى معنويـة $\alpha = 0.01$ عما إذا كان يوجد فرق معنوي بين وسيطي عمري النوعين من المصابيح.

٧ - البيانات التالية تمثل عدد الكيلومترات المقطوعة لكل جالون لثلاثة أنواع من الجازولين

النوع الأول: 20,31,24,33,23,24,28,16,19,26

النوع الثاني: 29,18,29,20,21,34,33,30,23,19

النوع الثالث: 19,31,16,31,33,26,28,28,25,30

استخدم اختبار كرسكال والس عند مستوى معنوية $\alpha = 0.05$ لاختبار فرض العدم H_0 بأنه لا يوجد اختلاف في متوسطات الكيلومترات للثلاثة أنـواع من الجازولين.

٨ - استخدم ثلاثة مستويات لمخدر على ثلاثة عينات من فئران التجارب وقيس الزمن بالثانية قبل النوم لكل عينة فكان كما يلي:

العينة الأولى الجرعة 0.5mg: 8.2,10.0,10.2,13.7,14.0,7.8,12.7,10.9

العينة الثانية الجرعة 1.0mg: 9.7,13.1,11.0,7.5,13.3,12.5,8.8,12.9
7.9,10.5

العينة الثالثة الجرعة 1.5mg: 12.0,7.2,8.0,9.4,11.3,9.0,11.5,8.5

استخدم اختبار كرسكال والس لاختبار مستوى الجرعات ليس له تـأثـير

معنوي بمستوى معنوية $\alpha = 0.05$.

٩ - أخذ رأي 50 شخصا لفرض ضريبة زيادة على الجازولين لاستخدامها في بناء طرق جديدة فكان المؤيد يأخذ الرمز F والمعارض يأخذ الرمز A ودونت النتائج كما يلي:

AAAFAFAAAAFFAAFAAAAFAAFF

AAFAAAAFAFFAAAAAFAAFAAAAF

استخدم اختبار الأشواط عند مستوى معنوية $\alpha = 0.05$ لاختبار ما إذا كان هذا الترتيب للرمز F والرمز A يأخذ الشكل العشوائي.

١٠ - اختبر ما إذا كان نوع القطعة المنتجة (معيبة وسليمة) من إحدى الآلات بصفة مستمرة للعينة التالية (حيث يرمز للقطعة المعيبة بالرمز d والسليمة بالرمز n) ينطبق عليها العشوائية حيث البيانات كما يلي:

nn ddd nnn dd nnnnnnn ddd nnnd nnnnn ddd

١١ - البيانات التالية تمثل عدد التلاميذ الغائبين في إحدى المدارس خلال 24 يوما متتالية اختبر ما إذا كانت هذه البيانات عشوائية أم لا عند مستوى معنوية $\alpha = 0.01$.

29, 25, 31, 28, 30, 28, 33, 31, 35, 29, 31, 33, 35, 28, 36

30, 33, 26, 30, 28, 32, 31, 38, 27

الجـداول

Tables

جدول رقم (١). التوزيع الطبيعي القياسي $\phi(z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Also, for z = 4.0, 5.0, and 6.0, the areas are 0.49997, 0.4999997, and 0.499999999.

جدول رقم (٢). توزيع t (ت)

d.f.	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	d.f.
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
inf.	1.282	1.645	1.960	2.326	2.576	inf.

† From Richard A. Johnson and Dean W. Wichern, *Applied Multivariate Statistical Analysis,* © 1982, p. 582. Adapted by permission of Prentice-Hall, Inc., Englewood Cliffs, N.J.

جدول رقم (٣). توزيع مربع كاي (χ^2)

d.f.	$\chi^2_{.995}$	$\chi^2_{.99}$	$\chi^2_{.975}$	$\chi^2_{.95}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$	d.f.
1	.0000393	.000157	.000982	.00393	3.841	5.024	6.635	7.879	1
2	.0100	.0201	.0506	.103	5.991	7.378	9.210	10.597	2
3	.0717	.115	.216	.352	7.815	9.348	11.345	12.838	3
4	.207	.297	.484	.711	9.488	11.143	13.277	14.860	4
5	.412	.554	.831	1.145	11.070	12.832	15.086	16.750	5
6	.676	.872	1.237	1.635	12.592	14.449	16.812	18.548	6
7	.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278	7
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955	8
9	1.735	2.088	2.700	3.325	13.919	19.023	21.666	23.589	9
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188	10
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757	11
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300	12
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819	13
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319	14
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801	15
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267	16
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718	17
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156	18
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582	19
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997	20
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401	21
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796	22
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181	23
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.558	24
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928	25
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290	26
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645	27
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993	28
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.336	29
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672	30

† Based on Table 8 of *Biometrika Tables for Statisticians, Volume I* (Cambridge University Press, 1954), by permission of the *Biometrika* trustees.

جدول رقم (٤) توزيع $F_{0.05}$

Degrees of freedom for numerator (columns); Degrees of freedom for denominator (rows)

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.29	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

† Reproduced from M. Merrington and C. M. Thompson, "Tables of percentage points of the inverted beta (F) distribution," *Biometrika*, vol. 33 (1943), by permission of the *Biometrika* trustees.

تابع جدول رقم (٤) توزيع $F_{0.01}$

Degrees of freedom for denominator

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4,052	5,000	5,403	5,625	5,764	5,859	5,928	5,982	6,023	6,056	6,106	6,157	6,209	6,235	6,261	6,287	6,313	6,339	6,366
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4	26.3	26.2	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.7	13.6	13.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.05	6.97	6.88
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.53	2.45	2.36	2.27	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

† Reproduced from M. Merrington and C. M. Thompson, "Tables of percentage points of the inverted beta (F) distribution," *Biometrika*, vol. 33 (1943), by permission of the *Biometrika* trustees.

جدول رقم (٥) الأرقام العشوائية

	00–04	05–09	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49
00	54463	22662	65905	70639	79365	67382	29085	69831	47058	08186
01	15389	85205	18850	39226	42249	90669	96325	23248	60933	26927
02	85941	40756	82414	02015	13858	78030	16269	65978	01385	15345
03	61149	69440	11286	88218	58925	03638	52862	62733	33451	77455
04	05219	81619	10651	67079	92511	59888	84502	72095	83463	75577
05	41417	98326	87719	92294	46614	50948	64886	20002	97365	30976
06	28357	94070	20652	35774	16249	75019	21145	05217	47286	76305
07	17783	00015	10806	83091	91530	36466	39981	62481	49177	75779
08	40950	84820	29881	85966	62800	70326	84740	62660	77379	90279
09	82995	64157	66164	41180	10089	41757	78258	96488	88629	37231
10	96754	17676	55659	44105	47361	34833	86679	23930	53249	27083
11	34357	88040	53364	71726	45690	66334	60332	22554	90600	71113
12	06318	37403	49927	57715	50423	67372	63116	48888	21505	80182
13	62111	52820	07243	79931	89292	84767	85693	73947	22278	11551
14	47534	09243	67879	00544	23410	12740	02540	54440	32949	13491
15	98614	75993	84460	62846	59844	14922	48730	73443	48167	34770
16	24856	03648	44898	09351	98795	18644	39765	71058	90368	44104
17	96887	12479	80621	66223	86085	78285	02432	53342	42846	94771
18	90801	21472	42815	77408	37390	76766	52615	32141	30268	18106
19	55165	77312	83666	36028	28420	70219	81369	41943	47366	41067
20	75884	12952	84318	95108	72305	64620	91318	89872	45375	85436
21	16777	37116	58550	42958	21460	43910	01175	87894	81378	10620
22	46230	43877	80207	88877	89380	32992	91380	03164	98656	59337
23	42902	66892	46134	01432	94710	23474	20423	60137	60609	13119
24	81007	00333	39693	28039	10154	95425	39220	19774	31782	49037
25	68089	01122	51111	72373	06902	74373	96199	97017	41273	21546
26	20411	67081	89950	16944	93054	87687	96693	87236	77054	33848
27	58212	13160	06468	15718	82627	76999	05999	58680	96739	63700
28	70577	42866	24969	61210	76046	67699	42054	12696	93758	03283
29	94522	74358	71659	62038	79643	79169	44741	05437	39038	13163
30	42626	86819	85651	88678	17401	03252	99547	32404	17918	62880
31	16051	33763	57194	16752	54450	19031	58580	47629	54132	60631
32	08244	27647	33851	44705	94211	46716	11738	55784	95374	72655
33	59497	04392	09419	89964	51211	04894	72882	17805	21896	83864
34	97155	13428	40293	09985	58434	01412	69124	82171	59058	82859
35	98409	66162	95763	47420	20792	61527	20441	39435	11859	41567
36	45476	84882	65109	96597	25930	66790	65706	61203	53634	22557
37	89300	69700	50741	30329	11658	23166	05400	66669	48708	03887
38	50051	95137	91631	66315	91428	12275	24816	68091	71710	33258
39	31753	85178	31310	89642	98364	02306	24617	09609	83942	22716
40	79152	53829	77250	20190	56535	18760	69942	77448	33278	48805
41	44560	38750	83635	56540	64900	42912	13953	79149	18710	68618
42	68328	83378	63369	71381	39564	05615	42451	64559	97501	65747
43	46939	38689	58625	08342	30459	85863	20781	09284	26333	91777
44	83544	86141	15707	96256	23068	13782	08467	89469	93842	55349
45	91621	00881	04900	54224	46177	55309	17852	27491	89415	23466
46	91896	67126	04151	03795	59077	11848	12630	98375	52068	60142
47	55751	62515	21108	80830	02263	29303	37204	96926	30506	09808
48	85156	87689	95493	88842	00664	55017	55539	17771	69448	87530
49	07521	56898	12236	60277	39102	62315	12239	07105	11844	01117

تابع جدول رقم (٥)

	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90–94	95–99
00	59391	58030	52098	82718	87024	82848	04190	96574	90464	29065
01	99567	76364	77204	04615	27062	96621	43918	01896	83991	51141
02	10363	97518	51400	25670	98342	61891	27101	37855	06235	33316
03	86859	19558	64432	16706	99612	59798	32803	67708	15297	28612
04	11258	24591	36863	55368	31721	94335	34936	02566	80972	08188
05	95068	88628	35911	14530	33020	80428	39936	31855	34334	64865
06	54463	47237	73800	91017	36239	71824	83671	39892	60518	37092
07	16874	62677	57412	13215	31389	62233	80827	73917	82802	84420
08	92494	63157	76593	91316	03505	72389	96363	52887	01087	66091
09	15669	56689	35682	40844	53256	81872	35213	09840	34471	74441
10	99116	75486	84989	23476	52967	67104	39495	39100	17217	74073
11	15696	10703	65178	90637	63110	17622	53988	71087	84148	11670
12	97720	15369	51269	69620	03388	13699	33423	67453	43269	56720
13	11666	13841	71681	98000	35979	39719	81899	07449	47985	46967
14	71628	73130	78783	75691	41632	09847	61547	18707	85489	69944
15	40501	51089	99943	91843	41995	88931	73631	69361	05375	15417
16	22518	55576	98215	82068	10798	86211	36584	67466	69373	40054
17	75112	30485	62173	02132	14878	92879	22281	16783	86352	00077
18	80327	02671	98191	84342	90813	49268	95441	15496	20168	09271
19	60251	45548	02146	05597	48228	81366	34598	72856	66762	17002
20	57430	82270	10421	00540	43648	75888	66049	21511	47676	33444
21	73528	39559	34434	88596	54086	71693	43132	14414	79949	85193
22	25991	65959	70769	64721	86413	33475	42740	06175	82758	66248
23	78388	16638	09134	59980	63806	48472	39318	35434	24057	74739
24	12477	09965	96657	57994	59439	76330	24596	77515	09577	91871
25	83266	32883	42451	15579	38155	29793	40914	65990	16255	17777
26	76970	80876	10237	39515	79152	74798	39357	09054	73579	92359
27	37074	65198	44785	68624	98336	84481	97610	78735	46703	98265
28	83712	06514	30101	78295	54656	85417	43189	60048	72781	72606
29	20287	56862	69727	94443	64936	08366	27227	05158	50326	59566
30	74261	32592	86538	27041	65172	85532	07571	80609	39285	65340
31	64081	49863	08478	96001	18888	14810	70545	89755	59064	07210
32	05617	75818	47750	67814	29575	10526	66192	44464	27058	40467
33	26793	74951	95466	74307	13330	42664	85515	20632	05497	33625
34	65988	72850	48737	54719	52056	01596	03845	35067	03134	70322
35	27366	42271	44300	73399	21105	03280	73457	43093	05192	48657
36	56760	10909	98147	34736	33863	95256	12731	66598	50771	83665
37	72880	43338	93643	58904	59543	23943	11231	83268	65938	81581
38	77888	38100	03062	58103	47961	83841	25878	23746	55903	44115
39	28440	07819	21580	51459	47971	29882	13990	29226	23608	15873
40	63525	94441	77033	12147	51054	49955	58312	76923	96071	05813
41	47606	93410	16359	89033	89696	47231	64498	31776	05383	39902
42	52669	45030	96279	14709	52372	87832	02735	50803	72744	88208
43	16738	60159	07425	62369	07515	82721	37875	71153	21315	00132
44	59348	11695	45751	15865	74739	05572	32688	20271	65128	14551
45	12900	71775	29845	60774	94924	21810	38636	33717	67598	82521
46	75086	23537	49939	33595	13484	97588	28617	17979	70749	35234
47	99495	51434	29181	09993	38190	42553	68922	52125	91077	40197
48	26075	31671	45386	36583	93459	48599	52022	41330	60651	91321
49	13636	93596	23377	51133	95126	61496	42474	45141	46660	42338

جدول رقم (٦) جدول دنكان {أقل مدى معنوي $q_a(k,\upsilon)$ عند $\alpha=0.05$ و $\alpha=0.1$}

Error df	Significance level	k' = number of means for range being tested													
		2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	.05	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
	.01	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	.05	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
	.01	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	.05	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
	.01	8.26	8.5	8.6	8.7	8.8	8.9	8.9	9.0	9.0	9.0	9.1	9.2	9.3	9.3
4	.05	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
	.01	6.51	6.8	6.9	7.0	7.1	7.1	7.2	7.2	7.3	7.3	7.4	7.4	7.5	7.5
5	.05	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
	.01	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.5	6.6	6.6	6.7	6.7	6.8
6	.05	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
	.01	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.0	6.1	6.2	6.2	6.3	6.3
7	.05	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61
	.01	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.8	5.8	5.9	5.9	6.0	6.0
8	.05	3.26	3.39	3.47	3.52	3.55	3.56	5.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56
	.01	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.5	5.6	5.7	5.7	5.8	5.8
9	.05	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52
	.01	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.4	5.5	5.5	5.6	5.7	5.7
10	.05	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.48
	.01	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.36	5.42	5.48	5.54	5.55
11	.05	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	4.46	3.46	3.46	3.46	3.47	3.48
	.01	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.24	5.28	5.34	5.38	5.39
12	.05	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.46	3.46	3.46	3.47	3.48
	.01	4.32	4.55	4.68	4.76	4.84	4.92	4.96	5.02	5.07	5.13	5.17	5.22	5.24	5.26
13	.05	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.45	3.46	3.46	3.47	3.47
	.01	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.04	5.08	5.13	5.14	5.15
14	.05	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47
	.01	4.21	4.42	4.55	4.63	4.70	4.78	4.83	4.87	4.91	4.96	5.00	5.04	5.06	5.07
15	.05	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44	3.45	3.46	3.47	3.47
	.01	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	4.90	4.94	4.97	4.99	5.00

تابع جدول رقم (٦) جدول دنكان

Error df	Significance level	k = number of means for range being tested												
		2	3	4	5	6	7	8	9	10	12	14	16	18	20
16	.05	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44	3.45	3.46	3.47	3.47
	.01	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.84	4.88	4.91	4.93	4.94
17	.05	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44	3.45	3.46	3.47	3.47
	.01	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75	4.80	4.83	4.86	4.88	4.89
18	.05	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43	3.45	3.46	3.47	3.47
	.01	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.76	4.79	4.82	4.84	4.85
19	.05	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43	3.44	3.46	3.47	3.47
	.01	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	4.67	4.72	4.76	4.79	4.81	4.82
20	.05	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43	3.44	3.46	3.46	3.47
	.01	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.69	4.73	4.76	4.78	4.79
22	.05	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42	3.44	3.45	3.46	3.47
	.01	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60	4.65	4.68	4.71	4.74	4.75
24	.05	2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41	3.44	3.45	3.46	3.47
	.01	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57	4.62	4.64	4.67	4.70	4.72
26	.05	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41	3.43	3.45	3.46	3.47
	.01	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53	4.58	4.62	4.65	4.67	4.69
28	.05	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40	3.43	3.45	3.46	3.47
	.01	3.91	4.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51	4.56	4.60	4.62	4.65	4.67
30	.05	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40	3.43	3.44	3.46	3.47
	.01	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.54	4.58	4.61	4.63	4.65
40	.05	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39	3.42	3.44	3.46	3.47
	.01	3.82	3.99	4.10	4.17	4.24	4.30	4.34	4.37	4.41	4.46	4.51	4.54	4.57	4.59
60	.05	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37	3.40	3.43	3.45	3.47
	.01	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.39	4.44	4.47	4.50	4.53
100	.05	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36	3.40	3.42	3.45	3.47
	.01	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29	4.35	4.38	4.42	4.45	4.48
∞	.05	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34	3.38	3.41	3.44	3.47
	.01	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.26	4.31	4.34	4.38	4.41

SOURCE: Abridged from D. B. Duncan, "Multiple range and multiple *F* tests," *Biometrics*, **11**: 1–42 (1955), with the permission of the editor and the author.

جدول رقم (٧). القيم الجدولية لاختبار ديونيت Dunnett (اتجاهين)*
$d_{\alpha}(t-1, \upsilon)$ عند $\alpha=0.05$

υ	t − 1 = Number of treatment means (excluding control)								
	1	2	3	4	5	6	7	8	9
5	2.57	3.03	3.29	3.48	3.62	3.73	3.82	3.90	3.97
6	2.45	2.86	3.10	3.26	3.39	3.49	3.57	3.64	3.71
7	2.36	2.75	2.97	3.12	3.24	3.33	3.41	3.47	3.53
8	2.31	2.67	2.88	3.02	3.13	3.22	3.29	3.35	3.41
9	2.26	2.61	2.81	2.95	3.05	3.14	3.20	3.26	3.32
10	2.23	2.57	2.76	2.89	2.99	3.07	3.14	3.19	3.24
11	2.20	2.53	2.72	2.84	2.94	3.02	3.08	3.14	3.19
12	2.18	2.50	2.68	2.81	2.90	2.98	3.04	3.09	3.14
13	2.16	2.48	2.65	2.78	2.87	2.94	3.00	3.06	3.10
14	2.14	2.46	2.63	2.75	2.84	2.91	2.97	3.02	3.07
15	2.13	2.44	2.61	2.73	2.82	2.89	2.95	3.00	3.04
16	2.12	2.42	2.59	2.71	2.80	2.87	2.92	2.97	3.02
17	2.11	2.41	2.58	2.69	2.78	2.85	2.90	2.95	3.00
18	2.10	2.40	2.56	2.68	2.76	2.83	2.89	2.94	2.98
19	2.09	2.39	2.55	2.66	2.75	2.81	2.87	2.92	2.96
20	2.09	2.38	2.54	2.65	2.73	2.80	2.86	2.90	2.95
24	2.06	2.35	2.51	2.61	2.70	2.76	2.81	2.86	2.90
30	2.04	2.32	2.47	2.58	2.66	2.72	2.77	2.82	2.86
40	2.02	2.29	2.44	2.54	2.62	2.68	2.73	2.77	2.81
60	2.00	2.27	2.41	2.51	2.58	2.64	2.69	2.73	2.77
120	1.98	2.24	2.38	2.47	2.55	2.60	2.65	2.69	2.73
∞	1.96	2.21	2.35	2.44	2.51	2.57	2.61	2.65	2.69

* تم الحصول على هذه الجداول من مقالة لديونيت بعنوان «جداول جديدة للمقارنات المتعددة مع العينة الضابطة» في مجلة البيومتركس، مجلد (٢٠)، عدد (٣)، سنة ١٩٦٤م وكذلك من مقالة أخرى لديونيت بعنوان «طرق المقارنات المتعددة لمقارنة بعض أنواع المعالجات مع العينة الضابطة» في الدورية الإحصائية الأمريكية المصاحبة، مجلد (٥٠)، سنة ١٩٥٥م.

تابع جدول رقم (٧). القيم الجدولية لاختبار ديونيت (اتجاهين)

$d_{\alpha}(t-1, \upsilon)$ عند $\alpha=0.01$

υ	t − 1 = number of treatment means (excluding control)								
	1	2	3	4	5	6	7	8	9
5	4.03	4.63	4.98	5.22	5.41	5.56	5.69	5.80	5.89
6	3.71	4.21	4.51	4.71	4.87	5.00	5.10	5.20	5.28
7	3.50	3.95	4.21	4.39	4.53	4.64	4.74	4.82	4.89
8	3.36	3.77	4.00	4.17	4.29	4.40	4.48	4.56	4.62
9	3.25	3.63	3.85	4.01	4.12	4.22	4.30	4.37	4.43
10	3.17	3.53	3.74	3.88	3.99	4.08	4.16	4.22	4.28
11	3.11	3.45	3.65	3.79	3.89	3.98	4.05	4.11	4.16
12	3.05	3.39	3.58	3.71	3.81	3.89	3.96	4.02	4.07
13	3.01	3.33	3.52	3.65	3.74	3.82	3.89	3.94	3.99
14	2.98	3.29	3.47	3.59	3.69	3.76	3.83	3.88	3.93
15	2.95	3.25	3.43	3.55	3.64	3.71	3.78	3.83	3.88
16	2.92	3.22	3.39	3.51	3.60	3.67	3.73	3.78	3.83
17	2.90	3.19	3.36	3.47	3.56	3.63	3.69	3.74	3.79
18	2.88	3.17	3.33	3.44	3.53	3.60	3.66	3.71	3.75
19	2.86	3.15	3.31	3.42	3.50	3.57	3.63	3.68	3.72
20	2.85	3.13	3.29	3.40	3.48	3.55	3.60	3.65	3.69
24	2.80	3.07	3.22	3.32	3.40	3.47	3.52	3.57	3.61
30	2.75	3.01	3.15	3.25	3.33	3.39	3.44	3.49	3.52
40	2.70	2.95	3.09	3.19	3.26	3.32	3.37	3.41	3.44
60	2.66	2.90	3.03	3.12	3.19	3.25	3.29	3.33	3.37
120	2.62	2.85	2.97	3.06	3.12	3.18	3.22	3.26	3.29
∞	2.58	2.79	2.92	3.00	3.06	3.11	3.15	3.19	3.22

تابع جدول رقم (٧). القيم الجدولية لاختبار ديونيت Dunnett (اتجاه واحد)
$d_{\alpha}(t-1, \upsilon)$ عند $\alpha=0.05$

υ	t − 1 = number of treatment means (excluding control)								
	1	2	3	4	5	6	7	8	9
5	2.02	2.44	2.68	2.85	2.98	3.08	3.16	3.24	3.30
6	1.94	2.34	2.56	2.71	2.83	2.92	3.00	3.07	3.12
7	1.89	2.27	2.48	2.62	2.73	2.82	2.89	2.95	3.01
8	1.86	2.22	2.42	2.55	2.66	2.74	2.81	2.87	2.92
9	1.83	2.18	2.37	2.50	2.60	2.68	2.75	2.81	2.86
10	1.81	2.15	2.34	2.47	2.56	2.64	2.70	2.76	2.81
11	1.80	2.13	2.31	2.44	2.53	2.60	2.67	2.72	2.77
12	1.78	2.11	2.29	2.41	2.50	2.58	2.64	2.69	2.74
13	1.77	2.09	2.27	2.39	2.48	2.55	2.61	2.66	2.71
14	1.76	2.08	2.25	2.37	2.46	2.53	2.59	2.64	2.69
15	1.75	2.07	2.24	2.36	2.44	2.51	2.57	2.62	2.67
16	1.75	2.06	2.23	2.34	2.43	2.50	2.56	2.61	2.65
17	1.74	2.05	2.22	2.33	2.42	2.49	2.54	2.59	2.64
18	1.73	2.04	2.21	2.32	2.41	2.48	2.53	2.58	2.62
19	1.73	2.03	2.20	2.31	2.40	2.47	2.52	2.57	2.61
20	1.72	2.03	2.19	2.30	2.39	2.46	2.51	2.56	2.60
24	1.71	2.01	2.17	2.28	2.36	2.43	2.48	2.53	2.57
30	1.70	1.99	2.15	2.25	2.33	2.40	2.45	2.50	2.54
40	1.68	1.97	2.13	2.23	2.31	2.37	2.42	2.47	2.51
60	1.67	1.95	2.10	2.21	2.28	2.35	2.39	2.44	2.48
120	1.66	1.93	2.08	2.18	2.26	2.32	2.37	2.41	2.45
∞	1.64	1.92	2.06	2.16	2.23	2.29	2.34	2.38	2.42

تابع جدول رقم (٧). القيم الجدولية لاختبار ديونيت **Dunnett** (اتجاه واحد)

$d_{\alpha}(t-1, \upsilon)$ عند $\alpha=0.01$

υ	$t-1$ = number of treatment means (excluding control)								
	1	2	3	4	5	6	7	8	9
5	3.37	3.90	4.21	4.43	4.60	4.73	4.85	4.94	5.03
6	3.14	3.61	3.88	4.07	4.21	4.33	4.43	4.51	4.59
7	3.00	3.42	3.66	3.83	3.96	4.07	4.15	4.23	4.30
8	2.90	3.29	3.51	3.67	3.79	3.88	3.96	4.03	4.09
9	2.82	3.19	3.40	3.55	3.66	3.75	3.82	3.89	3.94
10	2.76	3.11	3.31	3.45	3.56	3.64	3.71	3.78	3.83
11	2.72	3.06	3.25	3.38	3.48	3.56	3.63	3.69	3.74
12	2.68	3.01	3.19	3.32	3.42	3.50	3.56	3.62	3.67
13	2.65	2.97	3.15	3.27	3.37	3.44	3.51	3.56	3.61
14	2.62	2.94	3.11	3.23	3.32	3.40	3.46	3.51	3.56
15	2.60	2.91	3.08	3.20	3.29	3.36	3.42	3.47	3.52
16	2.58	2.88	3.05	3.17	3.26	3.33	3.39	3.44	3.48
17	2.57	2.86	3.03	3.14	3.23	3.30	3.36	3.41	3.45
18	2.55	2.84	3.01	3.12	3.21	3.27	3.33	3.38	3.42
19	2.54	2.83	2.99	3.10	3.18	3.25	3.31	3.36	3.40
20	2.53	2.81	2.97	3.08	3.17	3.23	3.29	3.34	3.38
24	2.49	2.77	2.92	3.03	3.11	3.17	3.22	3.27	3.31
30	2.46	2.72	2.87	2.97	3.05	3.11	3.16	3.21	3.24
40	2.42	2.68	2.82	2.92	2.99	3.05	3.10	3.14	3.18
60	2.39	2.64	2.78	2.87	2.94	3.00	3.04	3.08	3.12
120	2.36	2.60	2.73	2.82	2.89	2.94	2.99	3.03	3.06
∞	2.33	2.56	2.68	2.77	2.84	2.89	2.93	2.97	3.00

جدول رقم (٨). توزيع ذي الحدين b(n,p)

						p						
n	x	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
2	0	0.902	0.810	0.640	0.490	0.360	0.250	0.160	0.090	0.040	0.010	0.002
	1	0.095	0.180	0.320	0.420	0.480	0.500	0.480	0.420	0.320	0.180	0.095
	2	0.002	0.010	0.040	0.090	0.160	0.250	0.360	0.490	0.640	0.810	0.902
3	0	0.857	0.729	0.512	0.343	0.216	0.125	0.064	0.027	0.008	0.001	
	1	0.135	0.243	0.384	0.441	0.432	0.375	0.288	0.189	0.096	0.027	0.007
	2	0.007	0.027	0.096	0.189	0.288	0.375	0.432	0.441	0.384	0.243	0.135
	3		0.001	0.008	0.027	0.064	0.125	0.216	0.343	0.512	0.729	0.857
4	0	0.815	0.656	0.410	0.240	0.130	0.062	0.026	0.008	0.002		
	1	0.171	0.292	0.410	0.412	0.346	0.250	0.154	0.076	0.026	0.004	
	2	0.014	0.049	0.154	0.265	0.346	0.375	0.346	0.265	0.154	0.049	0.014
	3		0.004	0.026	0.076	0.154	0.250	0.346	0.412	0.410	0.292	0.171
	4			0.002	0.008	0.026	0.062	0.130	0.240	0.410	0.656	0.815
5	0	0.774	0.590	0.328	0.168	0.078	0.031	0.010	0.002			
	1	0.204	0.328	0.410	0.360	0.259	0.156	0.077	0.028	0.006		
	2	0.021	0.073	0.205	0.309	0.346	0.312	0.230	0.132	0.051	0.008	0.001
	3	0.001	0.008	0.051	0.132	0.230	0.312	0.346	0.309	0.205	0.073	0.021
	4			0.006	0.028	0.077	0.156	0.259	0.360	0.410	0.328	0.204
	5				0.002	0.010	0.031	0.078	0.168	0.328	0.590	0.774
6	0	0.735	0.531	0.262	0.118	0.047	0.016	0.004	0.001			
	1	0.232	0.354	0.393	0.303	0.187	0.094	0.037	0.010	0.002		
	2	0.031	0.098	0.246	0.324	0.311	0.234	0.138	0.060	0.015	0.001	
	3	0.002	0.015	0.082	0.185	0.276	0.312	0.276	0.185	0.082	0.015	0.002
	4		0.001	0.015	0.060	0.138	0.234	0.311	0.324	0.246	0.098	0.031
	5			0.002	0.010	0.037	0.094	0.187	0.303	0.393	0.354	0.232
	6				0.001	0.004	0.016	0.047	0.118	0.262	0.531	0.735
7	0	0.698	0.478	0.210	0.082	0.028	0.008	0.002				
	1	0.257	0.372	0.367	0.247	0.131	0.055	0.017	0.004			
	2	0.041	0.124	0.275	0.318	0.261	0.164	0.077	0.025	0.004		
	3	0.004	0.023	0.115	0.227	0.290	0.273	0.194	0.097	0.029	0.003	
	4		0.003	0.029	0.097	0.194	0.273	0.290	0.227	0.115	0.023	0.004
	5			0.004	0.025	0.077	0.164	0.261	0.318	0.275	0.124	0.041
	6				0.004	0.017	0.055	0.131	0.247	0.367	0.372	0.257
	7					0.002	0.008	0.028	0.082	0.210	0.478	0.698
8	0	0.663	0.430	0.168	0.058	0.017	0.004	0.001				
	1	0.279	0.383	0.336	0.198	0.090	0.031	0.008	0.001			
	2	0.051	0.149	0.294	0.296	0.209	0.109	0.041	0.010	0.001		
	3	0.005	0.033	0.147	0.254	0.279	0.219	0.124	0.047	0.009		
	4		0.005	0.046	0.136	0.232	0.273	0.232	0.136	0.046	0.005	
	5			0.009	0.047	0.124	0.219	0.279	0.254	0.147	0.033	0.005
	6			0.001	0.010	0.041	0.109	0.209	0.296	0.294	0.149	0.051
	7				0.001	0.008	0.031	0.090	0.198	0.336	0.383	0.279
	8					0.001	0.004	0.017	0.058	0.168	0.430	0.663

All values omitted in this table are 0.0005 or less.

تابع جدول رقم (٨). توزيع ذي الحدين b(n,p)

n	x	p = 0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
9	0	0.630	0.387	0.134	0.040	0.010	0.002					
	1	0.299	0.387	0.302	0.156	0.060	0.018	0.004				
	2	0.063	0.172	0.302	0.267	0.161	0.070	0.021	0.004			
	3	0.008	0.045	0.176	0.267	0.251	0.164	0.074	0.021	0.003		
	4	0.001	0.007	0.066	0.172	0.251	0.246	0.167	0.074	0.017	0.001	
	5		0.001	0.017	0.074	0.167	0.246	0.251	0.172	0.066	0.007	0.001
	6			0.003	0.021	0.074	0.164	0.251	0.267	0.176	0.045	0.008
	7				0.004	0.021	0.070	0.161	0.267	0.302	0.172	0.063
	8					0.004	0.018	0.060	0.156	0.302	0.387	0.299
	9						0.002	0.010	0.040	0.134	0.387	0.630
10	0	0.599	0.349	0.107	0.028	0.006	0.001					
	1	0.315	0.387	0.268	0.121	0.040	0.010	0.002				
	2	0.075	0.194	0.302	0.233	0.121	0.044	0.011	0.001			
	3	0.010	0.057	0.201	0.267	0.215	0.117	0.042	0.009	0.001		
	4	0.001	0.011	0.088	0.200	0.251	0.205	0.111	0.037	0.006		
	5		0.001	0.026	0.103	0.201	0.246	0.201	0.103	0.026	0.001	
	6			0.006	0.037	0.111	0.205	0.251	0.200	0.088	0.011	0.001
	7			0.001	0.009	0.042	0.117	0.215	0.267	0.201	0.057	0.010
	8				0.001	0.011	0.044	0.121	0.233	0.302	0.194	0.075
	9					0.002	0.010	0.040	0.121	0.268	0.387	0.315
	10						0.001	0.006	0.028	0.107	0.349	0.599
11	0	0.569	0.314	0.086	0.020	0.004						
	1	0.329	0.384	0.236	0.093	0.027	0.005	0.001				
	2	0.087	0.213	0.295	0.200	0.089	0.027	0.005	0.001			
	3	0.014	0.071	0.221	0.257	0.177	0.081	0.023	0.004			
	4	0.001	0.016	0.111	0.220	0.236	0.161	0.070	0.017	0.002		
	5		0.002	0.039	0.132	0.221	0.226	0.147	0.057	0.010		
	6			0.010	0.057	0.147	0.226	0.221	0.132	0.039	0.002	
	7			0.002	0.017	0.070	0.161	0.236	0.220	0.111	0.016	0.001
	8				0.004	0.023	0.081	0.177	0.257	0.221	0.071	0.014
	9				0.001	0.005	0.027	0.089	0.200	0.295	0.213	0.087
	10					0.001	0.005	0.027	0.093	0.236	0.384	0.329
	11							0.004	0.020	0.086	0.314	0.569
12	0	0.540	0.282	0.069	0.014	0.002						
	1	0.341	0.377	0.206	0.071	0.017	0.003					
	2	0.099	0.230	0.283	0.168	0.064	0.016	0.002				
	3	0.017	0.085	0.236	0.240	0.142	0.054	0.012	0.001			
	4	0.002	0.021	0.133	0.231	0.213	0.121	0.042	0.008	0.001		
	5		0.004	0.053	0.158	0.227	0.193	0.101	0.029	0.003		
	6			0.016	0.079	0.177	0.226	0.177	0.079	0.016		
	7			0.003	0.029	0.101	0.193	0.227	0.158	0.053	0.004	
	8			0.001	0.008	0.042	0.121	0.213	0.231	0.133	0.021	0.002
	9				0.001	0.012	0.054	0.142	0.240	0.236	0.085	0.017
	10					0.002	0.016	0.064	0.168	0.283	0.230	0.099
	11						0.003	0.017	0.071	0.206	0.377	0.341
	12							0.002	0.014	0.069	0.282	0.540

تابع جدول رقم (٨). توزيع ذي الحدين **b(n,p)**

		p										
n	*x*	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
13	0	0.513	0.254	0.055	0.010	0.001						
	1	0.351	0.367	0.179	0.054	0.011	0.002					
	2	0.111	0.245	0.268	0.139	0.045	0.010	0.001				
	3	0.021	0.100	0.246	0.218	0.111	0.035	0.006	0.001			
	4	0.003	0.028	0.154	0.234	0.184	0.087	0.024	0.003			
	5		0.006	0.069	0.180	0.221	0.157	0.066	0.014	0.001		
	6		0.001	0.023	0.103	0.197	0.209	0.131	0.044	0.006		
	7			0.006	0.044	0.131	0.209	0.197	0.103	0.023	0.001	
	8			0.001	0.014	0.066	0.157	0.221	0.180	0.069	0.006	
	9				0.003	0.024	0.087	0.184	0.234	0.154	0.028	0.003
	10				0.001	0.006	0.035	0.111	0.218	0.246	0.100	0.021
	11					0.001	0.010	0.045	0.139	0.268	0.245	0.111
	12						0.002	0.011	0.054	0.179	0.367	0.351
	13							0.001	0.010	0.055	0.254	0.513
14	0	0.488	0.229	0.044	0.007	0.001						
	1	0.359	0.356	0.154	0.041	0.007	0.001					
	2	0.123	0.257	0.250	0.113	0.032	0.006	0.001				
	3	0.026	0.114	0.250	0.194	0.085	0.022	0.003				
	4	0.004	0.035	0.172	0.229	0.155	0.061	0.014	0.001			
	5		0.008	0.086	0.196	0.207	0.122	0.041	0.007			
	6		0.001	0.032	0.126	0.207	0.183	0.092	0.023	0.002		
	7			0.009	0.062	0.157	0.209	0.157	0.062	0.009		
	8			0.002	0.023	0.092	0.183	0.207	0.126	0.032	0.001	
	9				0.007	0.041	0.122	0.207	0.196	0.086	0.008	
	10				0.001	0.014	0.061	0.155	0.229	0.172	0.035	0.004
	11					0.003	0.022	0.085	0.194	0.250	0.114	0.026
	12					0.001	0.006	0.032	0.113	0.250	0.257	0.123
	13						0.001	0.007	0.041	0.154	0.356	0.359
	14							0.001	0.007	0.044	0.229	0.488
15	0	0.463	0.206	0.035	0.005							
	1	0.366	0.343	0.132	0.031	0.005						
	2	0.135	0.267	0.231	0.092	0.022	0.003					
	3	0.031	0.129	0.250	0.170	0.063	0.014	0.002				
	4	0.005	0.043	0.188	0.219	0.127	0.042	0.007	0.001			
	5	0.001	0.010	0.103	0.206	0.186	0.092	0.024	0.003			
	6		0.002	0.043	0.147	0.207	0.153	0.061	0.012	0.001		
	7			0.014	0.081	0.177	0.196	0.118	0.035	0.003		
	8			0.003	0.035	0.118	0.196	0.177	0.081	0.014		
	9			0.001	0.012	0.061	0.153	0.207	0.147	0.043	0.002	
	10				0.003	0.024	0.092	0.186	0.206	0.103	0.010	0.001
	11				0.001	0.007	0.042	0.127	0.219	0.188	0.043	0.005
	12					0.002	0.014	0.063	0.170	0.250	0.129	0.031
	13						0.003	0.022	0.092	0.231	0.267	0.135
	14							0.005	0.031	0.132	0.343	0.366
	15								0.005	0.035	0.206	0.463

تابع جدول رقم (٨). توزيع ذي الحدين b(n,p)

n	x	0.05	0.1	0.2	0.3	p 0.4	0.5	0.6	0.7	0.8	0.9	0.95
16	0	0.440	0.185	0.028	0.003							
	1	0.371	0.329	0.113	0.023	0.003						
	2	0.146	0.275	0.211	0.073	0.015	0.002					
	3	0.036	0.142	0.246	0.146	0.047	0.009	0.001				
	4	0.006	0.051	0.200	0.204	0.101	0.028	0.004				
	5	0.001	0.014	0.120	0.210	0.162	0.067	0.014	0.001			
	6		0.003	0.055	0.165	0.198	0.122	0.039	0.006			
	7			0.020	0.101	0.189	0.175	0.084	0.019	0.001		
	8			0.006	0.049	0.142	0.196	0.142	0.049	0.006		
	9			0.001	0.019	0.084	0.175	0.189	0.101	0.020		
	10				0.006	0.039	0.122	0.198	0.165	0.055	0.003	
	11				0.001	0.014	0.067	0.162	0.210	0.120	0.014	0.001
	12					0.004	0.028	0.101	0.204	0.200	0.051	0.006
	13					0.001	0.009	0.047	0.146	0.246	0.142	0.036
	14						0.002	0.015	0.073	0.211	0.275	0.146
	15							0.003	0.023	0.113	0.329	0.371
	16								0.003	0.028	0.185	0.440
17	0	0.418	0.167	0.023	0.002							
	1	0.374	0.315	0.096	0.017	0.002						
	2	0.158	0.280	0.191	0.058	0.010	0.001					
	3	0.041	0.156	0.239	0.125	0.034	0.005					
	4	0.008	0.060	0.209	0.187	0.080	0.018	0.002				
	5	0.001	0.017	0.136	0.208	0.138	0.047	0.008	0.001			
	6		0.004	0.068	0.178	0.184	0.094	0.024	0.003			
	7		0.001	0.027	0.120	0.193	0.148	0.057	0.009			
	8			0.008	0.064	0.161	0.185	0.107	0.028	0.002		
	9			0.002	0.028	0.107	0.185	0.161	0.064	0.008		
	10				0.009	0.057	0.148	0.193	0.120	0.027	0.001	
	11				0.003	0.024	0.094	0.184	0.178	0.068	0.004	
	12				0.001	0.008	0.047	0.138	0.208	0.136	0.017	0.001
	13					0.002	0.018	0.080	0.187	0.209	0.060	0.008
	14						0.005	0.034	0.125	0.239	0.156	0.041
	15						0.001	0.010	0.058	0.191	0.280	0.158
	16							0.002	0.017	0.096	0.315	0.374
	17								0.002	0.023	0.167	0.418
18	0	0.397	0.150	0.018	0.002							
	1	0.376	0.300	0.081	0.013	0.001						
	2	0.168	0.284	0.172	0.046	0.007	0.001					
	3	0.047	0.168	0.230	0.105	0.025	0.003					
	4	0.009	0.070	0.215	0.168	0.061	0.012	0.001				
	5	0.001	0.022	0.151	0.202	0.115	0.033	0.004				
	6		0.005	0.082	0.187	0.166	0.071	0.015	0.001			
	7		0.001	0.035	0.138	0.189	0.121	0.037	0.005			
	8			0.012	0.081	0.173	0.167	0.077	0.015	0.001		
	9			0.003	0.039	0.128	0.185	0.128	0.039	0.003		
	10			0.001	0.015	0.077	0.167	0.173	0.081	0.012		
	11				0.005	0.037	0.121	0.189	0.138	0.035	0.001	

تابع جدول رقم (٨). توزيع ذي الحدين b(n,p)

						p						
n	*x*	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
	12				0.001	0.015	0.071	0.166	0.187	0.082	0.005	
	13					0.004	0.033	0.115	0.202	0.151	0.022	0.001
	14					0.001	0.012	0.061	0.168	0.215	0.070	0.009
	15						0.003	0.025	0.105	0.230	0.168	0.047
	16						0.001	0.007	0.046	0.172	0.284	0.168
	17							0.001	0.013	0.081	0.300	0.376
	18								0.002	0.018	0.150	0.397
19	0	0.377	0.135	0.014	0.001							
	1	0.377	0.285	0.068	0.009	0.001						
	2	0.179	0.285	0.154	0.036	0.005						
	3	0.053	0.180	0.218	0.087	0.017	0.002					
	4	0.011	0.080	0.218	0.149	0.047	0.007	0.001				
	5	0.002	0.027	0.164	0.192	0.093	0.022	0.002				
	6		0.007	0.095	0.192	0.145	0.052	0.008	0.001			
	7		0.001	0.044	0.153	0.180	0.096	0.024	0.002			
	8			0.017	0.098	0.180	0.144	0.053	0.008			
	9			0.005	0.051	0.146	0.176	0.098	0.022	0.001		
	10			0.001	0.022	0.098	0.176	0.146	0.051	0.005		
	11				0.008	0.053	0.144	0.180	0.098	0.017		
	12				0.002	0.024	0.096	0.180	0.153	0.044	0.001	
	13				0.001	0.008	0.052	0.145	0.192	0.095	0.007	
	14					0.002	0.022	0.093	0.192	0.164	0.027	0.002
	15					0.001	0.007	0.047	0.149	0.218	0.080	0.011
	16						0.002	0.017	0.087	0.218	0.180	0.053
	17							0.005	0.036	0.154	0.285	0.179
	18							0.001	0.009	0.068	0.285	0.377
	19								0.001	0.014	0.135	0.377
20	0	0.358	0.122	0.012	0.001							
	1	0.377	0.270	0.058	0.007							
	2	0.189	0.285	0.137	0.028	0.003						
	3	0.060	0.190	0.205	0.072	0.012	0.001					
	4	0.013	0.090	0.218	0.130	0.035	0.005					
	5	0.002	0.032	0.175	0.179	0.075	0.015	0.001				
	6		0.009	0.109	0.192	0.124	0.037	0.005				
	7		0.002	0.055	0.164	0.166	0.074	0.015	0.001			
	8			0.022	0.114	0.180	0.120	0.035	0.004			
	9			0.007	0.065	0.160	0.160	0.071	0.012			
	10			0.002	0.031	0.117	0.176	0.117	0.031	0.002		
	11				0.012	0.071	0.160	0.160	0.065	0.007		
	12				0.004	0.035	0.120	0.180	0.114	0.022		
	13				0.001	0.015	0.074	0.166	0.164	0.055	0.002	
	14					0.005	0.037	0.124	0.192	0.109	0.009	
	15					0.001	0.015	0.075	0.179	0.175	0.032	0.002
	16						0.005	0.035	0.130	0.218	0.090	0.013
	17						0.001	0.012	0.072	0.205	0.190	0.060
	18							0.003	0.028	0.137	0.285	0.189
	19								0.007	0.058	0.270	0.377
	20								0.001	0.012	0.122	0.358

المراجع

References

أولا: المراجع العربية

أبو صالح ، محمد صبحي وعوض ، عدنان محمد. *مقدمة في الإحصاء*. نيويورك : دار جون وايلي وأبنائه، ١٩٨٣م.

أبو عمه ،عبد الرحمن بن محمد سليمان؛ عبدالله ، أنور أحمد محمد وهندي ، محمود محمد ابراهيم. *الإحصاء التطبيقي* . الرياض : عمادة شؤون المكتبات، جامعة الملك سعود، ١٤١٠هـ (١٩٩٠م).

أحمد ، فاروق عبد العظيم؛ حسن، إمتثال محمد ؛ عزام ، عبد المرضي حامد و زغلول ، يحيى سعد. *مقدمة في الإحصاء* . إسكندرية : دار المطبوعات الجامعية، ١٩٨٣م.

الصياد ، جلال مصطفى وحبيب ، محمد الدسوقي. *مقدمة في الطرق الإحصائية*. جدة : دار عكاظ للطباعة والنشر، ١٩٩٠م .

الهانسي ، مختار محمد. *مقدمة في طرق التحليل الإحصائي* . بيروت : دار النهضة العربية للطباعة والنشر، ١٩٨٤م.

بري ، عدنان بن ماجد عبد الرحمن؛ هندي ، محمود محمد ابراهيم وعبدالله ، أنور أحمد محمد. *مبادىء الإحصاء والاحتمالات* . الرياض : عمادة شؤون المكتبات، جامعة الملك سعود، ١٤١٢هـ (١٩٩١م) .

سرحان ، أحمد عبـادة. *طرق التحليل الإحصائـي* . القاهرة : دار المعـارف، ١٩٦٥م.

كنجو ، أنيس اسماعيـل. *الإحصاء والاحتمـالات* . الرياض : عمادة شـؤون المكتبات، جامعة الملك سعود، ١٩٩٣م .

مصطفى ، مدني دسوقي. *مبادىء في نظرية الاحتمالات والإحصاء الرياضي* . القاهرة : دار النهضة، ١٩٧٩م.

هويل ، بول . ج . *المبادىء الأولية في الإحصاء*، الطبعة الثانية ، ترجمة : بدرية شوقي عبد الوهاب ومحمد كامل الشربيني . نيويورك : دار جون وايلي وأبنائه، ١٩٨٤م.

ثانيا: المراجع الأجنبية

Conover, W.J. *Practical Nonparametric Statistics,* 2nd ed. New York : John Wiley & Sons, 1980.

John, Freund E. *Modern Elementry Statistics,* 5th ed. Englewood Cliffs, New Jersy: Prentice hall International Editions, 1979.

Lapin, Lawrence. *Statistics Meaning and Methods.* New York : Harcourt Brace Jovanovich, Inc., 1980.

Mendenhall, Wiliam. *Introduction to Probability and Statistics.* North Scituate: Duxbury Press, 1980.

Montgomery, Douglas, C. *Design and Analysis of Experiments,* 2nd ed. New York: John Wiley & Sons, 1984.

Steel, Robert G.D. and Torrie, James H. *Principles and Procedures of Statistics a Biometrical Approach*, 2nd ed. New York : McGraw-Hill Book Company, 1980.

Weiss, N. and Hassett. *Introductory Statistics,* London : Adittison - Westey Publishing Company. Inc., 1982.

Zuwaylif, F.H. *General Applied Statistics,* 3rd ed. London : Adittiso Wesley Publishing Company Inc., 1979.

ثبت المصطلحات

- عربي – إنجليزي
- إنجليزي – عربي

أولاً: عربي – إنجليزي

أ

احتمالات الأحداث المستقلة	Probability of independent events
احتمال شرطي	Conditional probability
إحصاءات	Statistics
إحصاءة	Statistic
إحصاء وصفي	Descriptive statistics
اختبارات لا معلمية	Nonparametric test
اختبار الإشارة	Sign test
توكي	Tukeys' test
العشوائية	Test of randomness
كرسكال والس	Kruskal wllis test
مان وتني	Mann-whitney test
معلمي	Parametric test
ارتباط خطي	Linear correlation

استدلال	Inference
إحصائي	Statistical inference
أنحدار خطي	Linear regression
متعدد	Multiple linear regression
غير خطي	Nonlinear regression

ب

بيانات كميـــة	Quantitive data
مبوبة	Grouped data
نوعية (وصفية)	Qualitative data

ت

تأثير وسبب	Cause and effects
تباديل	Premutation
تباين	Variance
تجربة استطلاعية	Pilot experiment
تمهيدية	Preliminary experiment
حاسمة	Crucial experiment
ضابطة	Controlled experiment
عشوائية	Random experiment
تحليل التباين	Analysis of variance
تحيز	Bias
تشتت	Dispersion
تصميم تام التعشية	Complete randomized design
قطاعات كاملة العشوائية	Complete randomized block design

تعشية	Randomization
تقاطع	Intersection
تقدير بفترة	Interval estimation
تنبؤ	Predection
توزيع برنولي	Bernulli distribution
– ت	t - distribution
تراكمي	Comulative distribution
ذي الحدين	Binomial distribution
فوق الهندسي	Hypergeometric distribution
متصل	Continuous distribution
مربع كاي	Chi-square distribution
المعاينة	Sampling distributiion
منتظم متقطع	Discrete uniform distribution
منفصل (متقطع)	Discrete distribution

ج

جدول الاقتران	Contingency table
التوافق	Association table
جدولـــة	Tabulation

ح

حالات متماثلة (متساوية الفرصة)	Equaly likely cases
مواتية	Favourable cases
حجم منطقة الرفض	Size of rejection rejion
حدث	Event

حوادث تعتمد على الصدفة	Chance events
شاملة	Exhaustive events
مؤكدة (فراغ العينة)	Certain events
متنافية (منفصلة)	Mutually exclusive events

خ

خاصية التبديل	Commutative law
التجميع	Associative law
التوزيع	Distributive law
خطأ تجريبي	Experimental error
التحكم	Error control

د

دالة التوزيع التراكمي	Comulative distribution
الكتلة الاحتمالية	Probability mass function
الكثافة الاحتمالية	Probability density function
دقـــــة	Precision

ر

ربيعـــات	Quartiles

ط

طرق المقارنات المتعددة	Multiple comparison procedures
طريقة الصفة المميزة	Rule method
القائمـــة	Roster method
المربعات الصغرى	Least square method

ق

قاعدة الجمع	Addition rule
الضرب	Multiplication rule
قوانين دومورجان	De Morgan's laws

ك

كيفما اتفق	Haphazardness

م

مئينات	Precentiles
متضادات	Contrasts
متعامدة	Orthagnal contrasts
متغير تابع	Dependent variable
عشوائى	Random variable
متصل	Continuous random variable
مستقل (منفصل)	Independent random variable
متوسطات	Averages
مجموعات	Sets
مجموعة	Set
جزئية	Subset
خالية	Empty set
شاملة	Universal set
متممة	Complement set
محاولة برنولي	Bernulli trials

مدى جديد ومتعدد لدنكان	Duncans new multiple range
مشاهدات مصاحبة	Concomintant observations
مضلع تكراري	Polygon
متجمع صاعد	Increasing polygon
هابط	Decreasing polygon
معالجة	Treatment
معامل الاختلاف	Coefficient of variation
الاقتران	Coefficient of contingency
التحديد	Coefficient of determination
معاينة	Sampling
معلمة	Parameter
مقارنة	Comparsion
مقاييس التشتت	Measure of dispersion
النزعة المركزية	Measure of central tendency
منحنى تكراري	Frequency curve

ن

نظرية بيز	Bays theorem
تشبتشيف	Chebyshev's theorem
النهاية المركزية	Central limit theorem
نموذج احتمـــالي	Probabilistic model
مؤكـــد	Determinstic model

و

وحدة تجريبية	Experimental unit
المعاينة	Sampling unit
وسيط	Median

ثانيا: إنجليزي – عربي

A

Addition rule	قاعدة الجمع
Alternative hypothesis	فرض بديل
Analysis of variance	تحليل التباين
Arithmatic mean	وسط حسابي
Associative law	خاصية التجيمع
Averages	متوسطات

B

Bays theorem	نظرية بيز
Bernulli distributon	توزيع برنولي
trials	محاولة برنولي
Bias	تحيز
Binomial distribution	توزيع ذي الحدين (الثنائي)
Block	قطاع

C

Cause and effect	تأثير وسبب
Central limit theorem	نظرية النهاية المركزية
Certain events	حوادث مؤكدة
Chance events	حوادث معتمدة على الصدفة
Chebyshev's theorem	نظرية تشبتشيف

Chi-square distribution	توزيع مربع كاي
Coefficient of associations	معامل التوافق
of contingency	معامل الاقتران
of determination	معامل التجديد
of variation	معامل الاختلاف
Combinations	التوافيق
Commutative law	خاصية التبديل
Comparison	مقارنة
Complement set	مجموعة متممة
Complete	كاملـــة
Completely randomized block design	تصميم قطاعات كاملة التعشية
design	تصميم تام التعشية
Comulative distribution	توزيع تراكمي
Concomintant observations	مشاهدات مصاحبة
Conditional probability	احتمال شرطي
Contingency table	جدول الأقتران
Continuous distribution	توزيع متصل
Contrast	متضادة
Controlled expeiment	تجربة ضابطة
Correlation	ارتباط
Counting	العـــد (عد)
Crucial experiment	تجربة حاسمة

D

Decimles	عشيرات

مضلع متجمع هابط	Decreasing plogon
قوانين دي مورجان	De Morgan's laws
متغير تابع	Dependent variable
إحصاء وصفي	Descriptive statistics
نموذج مؤكد	Determenistic model
توزيع متقطع (منفصل)	Discrete distribution
متغيرعشوائي متقطع (منفصل)	random variable
توزيع منتظم متقطع	uniform distribution
تشتت	Dispersion
خاصية التوزيع	Distributive law
مدى متعدد وجديد لدنكان	Duncans new multiple range

E

مجموعة خالية	Empty set
حالات متماثلة (متساوية الفرصة)	Equaly likely cases
خطأ التحكم	Error control
حــدث	Event
أحداث	Events
حوادث شاملة	Exhaustive events
توقع	Expectation
تجربة	Experiment
تصميم تجريبي	Experimental design
خطأ تجريبي	error
وحدة تجريبية	unit

F

Favourable cases	حالات مواتية
Finite population	مجتمع محدود
Frequency curve	منحنى تكراري

G

Graphical representation	عرض الرسوم البيانية
Grouped data	بيانات مبوبة

H

Histogram	مدرج تكراري
Hypergeometric distribution	توزيع فوق هندسي
Hyphazardness	كيفما اتفق

I

Increasing polygons	مضلع متجمع صاعد
Independent events	استقلال الحوادث
Independent variable	متغير عشوائي مستقل
Inference	استدلال
Infinite population	مجتمع غير محدود
Input variable	متغير المدخل
Intersection	تقاطع

K

Kruskal wallis test	اختبار كرسكال والس

تفلطح	Kurtosis

L

أقل فرق معنوي	Least significant difference (LSD)
طريقة المربعات الصغرى	square method
مستوى المعنوية	Level of significance
ارتباط خطي	Linear correlation

M

اختبار مان وتني	Mann-whitney test
مقاييس النزعة المركزية	Measure of central tendency
مقاييس التشتت	of dispersion
وسيـــط	Median
منوال	Mode
طرق المقارنات المتعددة	Multiple comparison procedures
انحدار خطي متعدد	linear regression
حوادث متنافية (منفصلة)	Mutually exclusive events

N

فرض لا معلمي	Non parametric hypothesis
اختبارات اللا معلمية	parametric test
توزيع طبيعي	Normal distributions
فرض العدم	Null hypothesis

O

Orthognal contrasts	متضادات متعامدة

P

Parameter	معلمـــة
Parametric hypothesis	فروض حول معالم المجتمع
tests	اختبارات معلمية
Percentiles	مئينات
Pilot experiment	تجربة استطلاعية
Point estimation	تقدير بنقطة
Poisson distribution	توزيع بواسون
Polygon	مضلع تكراري
Population	مجتمع
Precision	دقـــة
Predection	تنبؤ
Preliminary experiment	تجربة تمهيدية
Premutation	تباديل
Probability axioms	مسلمات الاحتمال
density function	دالة كثافة احتمالية
distribution	توزيع احتمالي
mass function	دالة الكتلة الاحتمالية
of independent events	احتمالات الأحداث المستقلة

Qualitative data	بيانات نوعية
Quantative data	بيانات كمية
Quartiles	ربيعات
Questionnaire	استمارة إحصائية

R

Random experiment	تجربة عشوائية
Randomization	تعشيـــة
Random variable	متغير عشوائي
Range	مدى
Replication	مكرارات
Roster method	طريقة القائمة
Rule of multiplication	طريقة الصفة المميزة

S

Sample	عينة
space	فراغ العينة
Sampling distribution	توزيع المعاينة
of proportions	توزيع المعاينة للنسب
unit	وحدة المعاينة
Scatter diagram	مخطط انتشاري
Screening experiment	تجربة انتقائية
Semi interquartile range	نصف المدى الربيعي
Sets	مجموعات
Sighting experiment	تجربة تخطيطية

Sign test	اختبار الإشارة
Simple random sample	عينة عشوائية بسيطة
Size of rejection region	حجم منطقة الرفض
Standard deviation	انحراف معياري
normal distribution	توزيع طبيعي قياسي
Statistic	إحصاءة
Statistical hypothesis	فرض إحصائي
inference	استدلال إحصائي
population	مجتمع إحصائي
sampling	معاينة إحصائية
Statistics	إحصاءات

T

Tabulation	جدولة
T - distribution	توزيع ت
Test of randomness	اختبار العشوائية (الأشواط)
Treatment	معالجة
Tukey's procedure	اختبار توكي

U

Universal set	مجموعة شاملة

V

Variable	متغير
Variance	تباين

W

Weighted mean	وسط مرجح (موزون)

كشاف الموضوعات

ت

ن

و

ردمك :٣-٦٩٥-٠٥-٩٩٦٠
ISBN:9960-05-695-3